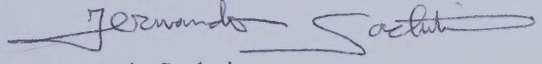
Dr. Smith,

In appreciation for your time examining and attending my thesis defence, I would like to give you a copy of my thesis for your records. Your time and effort was greatly appreciated.

Thank you,



Fernando Sacluti

University of Alberta

Library Release Form

Name of Author: Fernando R. Sacluti

Title of Thesis: Modeling Water Distribution Pipe Failures Using Artificial Neural Networks

Degree: Master of Science

Year this Degree Granted: 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

_____

Fernando R. Sacluti
20 Finch Crescent
St. Albert, AB
T8N 1Y6

Date: _____

University of Alberta

MODELING WATER DISTRIBUTION PIPE FAILURES USING ARTIFICIAL NEURAL NETWORKS

by

Fernando R. Sacluti

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Environmental Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta

Spring 1999

DEDICATION

This work is dedicated to my family and friends for their support, and especially to Anjum Mullick for all of her encouragement, support and tolerance during the course of this thesis.

ABSTRACT

Repair and replacement of cast iron water mains is a significant issue for many water utilities. Having the ability to predict the frequency of water main breaks gives utilities improved information for decision-making. In this study Artificial Neural Network (ANN) methodology was used to predict pipe break frequency, using historical pipe break data from a city subdivision.

The application of this modeling methodology is ideal due to its ability to use readily available meteorological and operational data. Due to the open-system nature of this study, extra care was taken to ensure use of reliable data. Having developed proven results, the developed ANN models could then be used to infer factors affecting pipe breaks, and to develop mitigation techniques.

ACKNOWLEDGMENTS

I would like to thank Aqualta for their funding of the thesis research as well as for supplying specific information for the water main breaks and operating parameters. Thanks should also be given to the Department of Meteorology and the Department of Civil and Environmental Engineering at the University of Alberta for their help in the research and development of the Artificial Neural Networks model. I would like to especially thank Dr. Steve Stanley and Mr. Qing Zhang for their help and guidance throughout the entire process.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1

1.1 Problem Statement 2

1.2 Water Main Replacement Overview 3

2.0 BACKGROUND INFORMATION 5

2.1 Pipe Failure Modes 5

2.2 Failure Mechanisms 7

2.2.1 Frost Heave 8

2.2.2 Soil-Pipeline Interaction 10

2.2.3 Pipe-Wall Temperature Gradients 10

2.2.4 Corrosion 11

2.2.5 Other 11

2.3 Parameters That Cause/Influence Breaks 12

2.4 Artificial Neural Networks Overview 32

2.4.1 Background Information and Application 32

2.4.2 ANN Model Development Process 34

2.4.3 ANN Model Structure 35

2.4.4 ANN Model Learning Mechanics 47

2.4.5 ANN Data Set Presentation 48

3.0 METHODOLOGY 49

3.1 Data Collection 49

3.1.1 Parameter Collection and Analysis 50

3.1.2 Open-Domain Problems 76

3.1.3 Limitations 78

3.2 Modeling Methodology 80

3.2.1 Model Progression *83*

3.2.2 Data Set Manipulation *86*

3.2.3 Data Point Manipulation: Selection/Elimination Protocol *88*

3.2.4 Input Parameter Selection: Methods Tried, Results *91*

4.0 RESULTS 95

4.1 Evaluation Criteria 95

4.1.1 R^2 Statistic *96*

4.1.2 Trends Prediction *97*

4.1.3 Model Simplicity *97*

4.2 Best Models (Cumulative Results) 97

4.3 Event Prediction Models 100

5.0 DISCUSSION 102

5.1 Sensitivity Analysis 102

5.2 Apparent Influential and Causal Factors 113

5.3 Model Capabilities, Limitations, and Uses 117

6.0 CONCLUSIONS AND RECOMMENDATIONS 120

REFERENCES 124

APPENDIX A. SAMPLE MODEL INPUT DATA, MODEL A 132

APPENDIX B. SAMPLE MODEL INPUT DATA, MODEL B. 139

LIST OF TABLES

Page

TABLE 1. COMMON RANGES OF RESISTIVITY FOR SOILS. 24

TABLE 2. SUMMARY OF INFLUENCING AND CAUSAL FACTORS. 31

TABLE 3. PIPE BREAKS SORTED BY DIAMETER, 1985-1991. 58

TABLE 4. SUMMARY OF POTENTIAL INPUT PARAMETERS. 75

TABLE 5. PROSPECTIVE MODEL A: MODEL SPECIFICS. 99

TABLE 6. PROSPECTIVE MODEL B: MODEL SPECIFICS. 99

LIST OF FIGURES

Page

FIGURE 1. CIRCULAR AND LONGITUDINAL SPLIT FAILURE MODES OF WATER MAINS 6

FIGURE 2. DIFFERENTIAL FROST HEAVE EFFECTS ON PIPELINES. 9

FIGURE 3. EFFECTS OF AIR POCKETS ON OPERATING PRESSURES. 16

FIGURE 4. WHIPLASH CURVE FOR GROUND-TO-AIR HEAT TRANSFER 28

FIGURE 5. TYPICAL ANN BACKPROPAGATION STRUCTURE. 36

FIGURE 6. LOGISTIC ACTIVATION FUNCTION 38

FIGURE 7. TANH ACTIVATION FUNCTION. 39

FIGURE 8. LINEAR ACTIVATION FUNCTION 40

FIGURE 9. GAUSSIAN ACTIVATION FUNCTION 41

FIGURE 10. GAUSSIAN COMPLEMENT ACTIVATION FUNCTION 42

FIGURE 11. STANDARD CONNECTION NETWORK STRUCTURE. 43

FIGURE 12. JUMP CONNECTION NETWORK STRUCTURES. 44

FIGURE 13. RECURRENT NETWORK STRUCTURES 44

FIGURE 14. WARD NETWORK STRUCTURES. 45

FIGURE 15. GENERAL MAP OF EDMONTON WATER DISTRIBUTION SYSTEM 53

FIGURE 16. CALDER SUB-DIVISION STUDY AREA 53

FIGURE 17. MONTHLY TRANSVERSE FAILURES, 1972-1994 55

FIGURE 18. MONTHLY DIAGONAL FAILURES, 1972-1994. 56

FIGURE 19. MONTHLY LONGITUDINAL FAILURES, 1972-1994. 56

FIGURE 20. MONTHLY BLOWOUT FAILURES, 1972-1994. 57

FIGURE 21. MONTHLY CLAMP FAILURES, 1972-1994. 57

FIGURE 22. AVERAGE DAILY TEMPERATURES, 1985 61

FIGURE 23. FIRST DIFFERENCE OF AVERAGE AIR TEMPERATURES, 1985-1986. 61

FIGURE 24. SECOND DIFFERENCE OF AVERAGE AIR TEMPERATURES, 1985-1986. 62

FIGURE 25. SEVEN-DAY AIR TEMPERATURE DIFFERENCE, 1985. 62

FIGURE 26. GRAPHICAL ANALYSIS OF ROSSLYN RESERVOIR AVERAGE WATER TEMPERATURE, 1985-1991. 66

FIGURE 27. SEVEN DAY AVERAGE WATER TEMPERATURE (ROSSDALE WATER TREATMENT PLANT) 1985-1991. 67

FIGURE 28. TOTAL MONTHLY BREAKS, 1985-1991. 71

FIGURE 29. PREVIOUS YEAR TOTAL PIPE BREAKS (MOVING TOTAL), 1985-1991. 74

FIGURE 30. EVENT PREDICTION BEFORE TRAINING WITH NON-BREAK DATA. 90

FIGURE 31. EVENT PREDICTION AFTER "POINT A" DATA POINT ADDITION 91

FIGURE 32. PROSPECTIVE MODEL A. 98

FIGURE 33. PROSPECTIVE MODEL B. 99

FIGURE 34. TYPICAL EVENT PREDICTION RESULTS 101

FIGURE 35. MODEL A SENSITIVITY ANALYSIS: 7-DAY AVERAGE AIR TEMPERATURE. 104

FIGURE 36. MODEL A SENSITIVITY ANALYSIS: 7-DAY AVERAGE WATER TEMPERATURE. 106

FIGURE 37. MODEL A SENSITIVITY ANALYSIS: PIPE-WALL TEMPERATURE DIFFERENTIAL. 107

FIGURE 38. MODEL A SENSITIVITY ANALYSIS: HISTORICAL 1-YEAR BREAK FREQUENCY. 108

FIGURE 39. MODEL B SENSITIVITY ANALYSIS: AIR TEMPERATURE. 109

FIGURE 40. MODEL B SENSITIVITY ANALYSIS: 7-DAY WATER TEMPERATURE. 110

FIGURE 41. MODEL B SENSITIVITY ANALYSIS: PIPE-WALL TEMPERATURE DIFFERENTIAL. 111

FIGURE 42. MODEL B SENSITIVITY ANALYSIS: SPATIAL CLUSTER INDEX. 112

FIGURE 43. MODEL B SENSITIVITY ANALYSIS: HISTORICAL 1-YEAR BREAK FREQUENCY. 113

ABBREVIATIONS

°C	degrees Celsius
°F	degrees Fahrenheit
AI	Artificial Intelligence
ANN	Artificial Neural Network
D	external pipe diameter
GRNN	General Regression Neural Network
G_s	soil shear modulus
H	bury depth of water mains from surface to the centre line of the pipe
K_0	coefficient of active resistance at rest
km	kilometre
km^2	squared kilometres
k_s	reaction modulus
mm	millimetre
mV	millivolt
pH	-log $[H^+]$

PNN	Probabilistic Neural Network
R^2	R-squared
s_u	undrained shear strength of clay
t	air temperature (°C)
u_y	displacement required to develop ultimate axial resistance
ν_s	Poisson's Ratio of soil
α	adhesion coefficient
δ	frictional angle between the pipe material and surrounding backfill
γ_s	submerged unit weight of soil

1.0 INTRODUCTION

In many urban areas, the frequency of cast iron water main pipe breaks has increased with time. The unavoidable question of how to deal with aging water distribution infrastructure has developed into issues of total costs associated with repair of pipe breaks (and consequent interruptions in water delivery service), and replacement of deteriorated water mains. These pipe breaks are viewed with greater seriousness than in previous times by the water utilities, from both an economic and a customer relation's standpoint. As a result, many utilities have sought methods for predicting these events to allow for proactive replacement of the deteriorated pipes, thus reducing the burden of emergency repairs. A secondary benefit of developing a rational model is to determine the prevailing nature of these failures, allowing for pipe break mitigative techniques where pipe replacement is not warranted.

The City of Edmonton's water utility, Aqualta, has sponsored the investigation of the use of Artificial Neural Networks (ANN) modeling to assist in the prediction of cast iron pipe break trends for city subdivisions. The purpose of the ANN model is to identify areas that will have a higher cumulative probability of cast iron distribution pipe failures, which necessitates the replacement of these problem water mains. To easily facilitate the use of the ANN models by the water utility, the information used for input parameters must be readily accessible due to the need for large quantities of historical data. Because there is a lack of available algorithms which describe the pipe break process, the Artificial Neural Networks modeling methodology is recommended.

1.1 Problem Statement

The purpose of this study was to investigate the feasibility of developing an Artificial Neural Network model to be used as a general screening tool. The expectation of this exercise was to confirm the potential utility of using the ANN methodology for modeling cast iron water pipe failures, accurately predicting the number of pipe breaks within a defined area for the purpose of determining the area's pipe break density. This information would be used as a criterion for the Cast Iron Renewal Program. This program advocates proactive replacement of the pipes in areas with the highest break frequency. The Calder subdivision within the City of Edmonton was chosen as the study area. For this study, the scope of the modeling was limited to 150 mm cast iron pipe. This was done largely due to availability of data.

Development of ANN models using this type of methodology requires the use of historically collected data. This data must be meaningful and easily accessible to prove model credibility and permit implementation of the model. Therefore, effectiveness of these models hinges on obtaining appropriate data. Having accomplished this collection task, the finished models would intrinsically capture the cause-effect logic of the pipe failure mechanisms. This would allow inferences to be made of the predominating pipe failure modes.

The study undertaken is of particular interest to cities in cold weather climates. Given that the majority of pipe failures causes are related to cold weather, the study is particular to areas where winter seasons involve sub-zero temperatures.

1.2 Water Main Replacement Overview

Aqualta's Engineering Department (formerly the City of Edmonton Water Branch, Water Network Engineering Section) is responsible for the maintenance of the water distribution infrastructure. Its water main replacement program is appropriately named the Cast Iron Renewal Program. The program's operating philosophy is based on the concept that the area with the highest cast iron pipe break density represents the greatest threat to service disruption, and therefore is the greatest priority for water main replacement. Therefore, the cost effectiveness of a selected renewal project is directly proportional to its failure frequency (i.e. replacement of high failure frequency mains will remove more potential failures from the system per dollar spent).

Studies show that a high percentage of the failures in the cast iron system tend to occur in a relatively small percentage (by area) of the city's water distribution system. These high frequency areas are normally designated for renewal based on a predetermined "critical failure frequency", in an effort to remove what is deemed to be the worst pipes from the system. This "critical failure frequency" is defined as a point above which it is more economical to replace the pipe, but below which it is more economical to repair pipe. In the past several years this critical frequency has been set at 5.0 failures/km/year

(O'Farrell, 1995), taken over a five year moving-average. The idea behind the process is to identify areas with consistently high historical break frequencies, and to replace these potential problem areas pro-actively. Funds are therefore allocated to replace a certain percentage of these areas per year.

The program has improved the service level by reducing the overall annual failure frequency from slightly over 1.0 failures/km/year in the mid-1980's, to 0.8 failures/km/year, ten years later—a 20 percent improvement.

With the program progressing towards its goal of replacing all pipe break densities equal to or greater than the five-year, moving average break density, new methods of maintaining flexibility in decision-making are required. The ability to accurately predict actual pipe break occurrences based on existing data is therefore advantageous.

2.0 BACKGROUND INFORMATION

2.1 Pipe Failure Modes

Statistics describing the performance of water mains are typically expressed in terms of a frequency of breaks. Breaks are defined as events leading to the disruption of water main service per kilometer per year.

Goulter and Kazemi presented two papers describing a study of the break frequency for Winnipeg, Manitoba [(Goulter and Kazemi, 1988) and (Goulter and Kazemi, 1989)]. This study is considered to be typical of many cities in North America and, therefore, will be used as a reference point.

Studies in several Canadian and U.S. cities have shown that water main breaks can occur in various modes of failure. These failure types include:

1. Circumferential failures (also referred to as circular or transverse);
2. Longitudinal split failures (includes diagonal);
3. Pinhole failures due to corrosion;
4. Pipe joint leaks (including fitting leaks) and;
5. Clamp failures.

These most common types of failures, circumferential and longitudinal split failures, are illustrated in Figure 1.

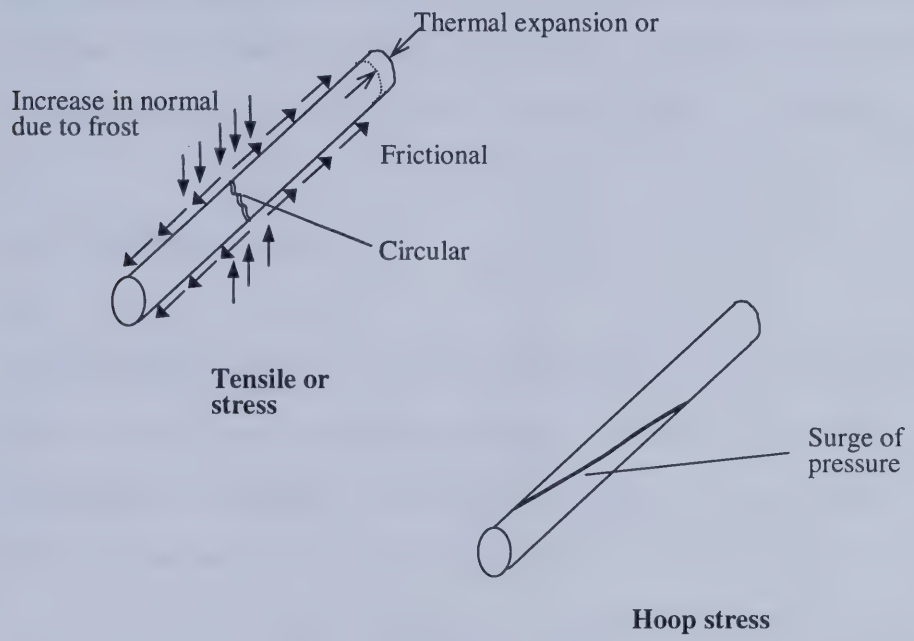


Figure 1. Circular and longitudinal split failure modes of water mains Adapted from (Rajani et al., 1996)

Although modal failure statistics vary per city, an average of 70 percent of water main failures are circumferential failures, with the remaining 30 percent being shared by the other types of failures (Rajani et al., 1996).

Circumferential failures are caused by longitudinal tensile stresses or from flexural (bending) stress. Longitudinal split failures are the result of circumferential (hoop) stress. Corrosion can also be the direct or indirect cause of pipe failures. These types of failures may be the result of a "blowout" (whereby a surge pressure causes the corrosion-

thinned wall to fail), or may be caused by complete wall corrosion. Joint failures can be the result of pipe jacking (heaving) or mis-aligned connections (Milligan, 1995). Clamp failures occur when clamps (used to repair previous pipe failures) themselves fail. In some cases, a combination of the above failures can occur. Therefore, these failures are usually the result of physical characteristics and environmental factors interacting.

2.2 Failure Mechanisms

Theories explaining the failure mechanisms of cast iron water mains in cold climate regions have advanced over the past two decades. Pipe failure modes have been identified, and the mechanisms thought to cause such failures have been studied. The mechanisms that predominate such discussions revolve around frost heave, soil-pipeline interaction, pipeline operating conditions and corrosion (both internal and external of the pipe).

The above mechanisms have been identified as likely causes for the different modes of failure of water mains. Circumferential pipe failures may be caused by excessive flexural or axial stresses (Habibian, 1994). Flexural stresses are thought to be the result of frost heave mechanisms (differential heave, beam-type loading). Axial stresses are caused by soil-pipeline frictional resistance opposing pipe shrinkage (brought on by sudden, extreme temperature drops). Longitudinal failures are thought to be caused by a high temperature gradient across the pipe wall, generating high hoop stresses (Habibian, 1994). Corrosion failures are thought to contribute both directly or indirectly to the

majority of all water main failures (Kettler and Goulter, 1985), and both internal and external pipe corrosion has been examined (Morris Jr., 1967). Pipe joint failures are also thought to be caused by frost heaving, or due to improper installation (Milligan, 1995).

2.2.1 Frost Heave

The frost heave mechanism has been a popular topic with regards to many types of structures. Almost any underground structure requires the consideration of frost heave effects that may displace portions or the entire underground structure. Frost heave is defined as the vertical expansion of soils caused by freezing of the soil and ice lens formation.

Differential heave causes sections of pipe to experience non-uniform displacements, and this differential results in forceful flexural stresses (Figure 2). Uniform heaving may also prove to be a problem under certain circumstances where pipe joints are not subject to movement. Under this scenario, the pipe experiences stresses similar to a simple beam loading, in which case the pipe will experience bending stresses. Failure of pipe joints may be the result of the frost heave process (i.e. pipe jacking) or due to illicit connections (Milligan, 1995). This may be a function of the type of connection, and the type of fill material used between joints.

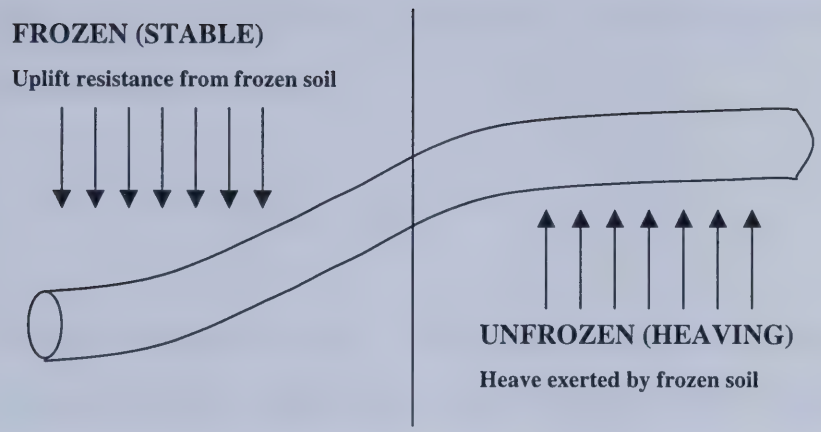


Figure 2. Differential frost heave effects on pipelines.
Adapted from (Nixon, 1994)

The principles of frost heave mechanics are well known in theory. Conditions for frost heave require the following (Anderson et al., 1984):

1. The presence of a frost susceptible soil;
2. The presence of a sufficient water source, whether it is capillary or a ground water source (for lens formation) and;
3. A ground temperature below zero degrees Celsius. Some argue that it is the change in temperature that is more important (Bahmanyar and Edil, 1983); others have argued that it is the difference between the prevailing temperature from the average temperature (Bates et al., 1996).

With all of the above factors present, there is the potential for damage due to frost heave. The propensity for heave of a soil under freezing conditions is affected by properties

such as grain size, rate of freezing, the availability of water, and by applied loads (Konrad, 1987).

2.2.2 Soil-Pipeline Interaction

Soil-pipeline interactions are also a possible cause of pipe failures. The resistance of the soil-to-pipeline union is important because the shear strength of the interaction can affect the degree of mobility of the pipeline and hence its ability to displace. In cold temperatures, the bond between the soil and pipe indicates the amount of restraint the pipe is allowed to shrink axially. A high soil-pipeline interaction will not allow the pipe to contract, and consequently the axial stress in the pipe will increase. It is also possible that a strong bond between the iron pipe and soil will cause excessive soil-pipe interface shear that may cause abrasion of the pipe coating. This abrasion may lead to premature corrosion of the pipe exterior (Yen et al., 1981).

2.2.3 Pipe-Wall Temperature Gradients

For longitudinal failures, a suspected failure mechanism is the high temperature gradient occurring across the pipe wall. If the temperature difference of the transported water and surrounding soil is significant, this temperature gradient can lead to unusually high hoop stresses, subsequently leading to failure (possibly due to a water pressure surge) (Habibian, 1994). Longitudinal failures may also occur in combination with the weakening of the pipe wall due to corrosion, at the weakest portion of the main wall.

Another possible cause of longitudinal failure is due to a crushing load. This usually occurs in the larger diameter pipes (O'Day, 1982).

2.2.4 Corrosion

Corrosion failure is directly or indirectly associated with the reduction in pipe wall thickness. These failures may occur as the result of internal or external corrosion. Possible types of internal corrosion include: bacteriological (Lutey and Mason, 1994); chemical corrosion occurring as internal pitting (Morris Jr., 1967) and (Quraishi and Al-Amry, 1992); or galvanic action (Morris Jr., 1967), although this type is less common. External corrosion may be caused by: galvanic action (Morris Jr., 1967); electrolytic oxidation due to low pH or stray currents (O'Day, 1982), or; bacterial, as sulfate reduction (Morris Jr., 1967). Potential sources of stray direct current in Edmonton may include electric railways (transit system and the Light Rail Transit system) and industrial equipment. The corrosion-weakened wall may fail by pin-hole failure, or may result from combination with one of the above mentioned failures ("blowout" or longitudinal failures).

2.2.5 Other

There are other factors that must be considered as causes for pipe failures. These causes are unpredictable circumstances that must be accounted for, or eliminated in the ANN models. Special phenomena such as spatial and temporal clustering of pipe breaks must

also be investigated. All of these factors should be examined for ANN models, such that the models will not confuse these types of events with frost heave, soil-pipeline interaction, pipe wall temperature gradient, and corrosion failure mechanisms.

Discussion of the phenomena of spatial and temporal clustering of pipe failures must be studied. The potential of vehicle loading must also be examined. Pipe age and distance from facilities may be considered, but some studies argue that pipe age is not a major determinant of water main break rates (O'Day, 1982). Instantaneous pressure surges causing water hammer and sudden pressure changes are examples of unpredictable events (which may cause multiple failures). These above types of failures must be considered, and then either accounted for or eliminated before modeling can proceed.

Having analyzed the type of failures and then summarizing the failure mechanisms, the Artificial Neural Network methodology must use reasonable model input parameters, which will allow it to characterize the cause-and-effect relationship of pipe failure mechanisms. Potential input parameters, based on the failure mechanisms, are outlined in the following section, with justification for their significance.

2.3 Parameters That Cause/Influence Breaks

In this section a discussion of potential causal and influencing factors effecting pipe break failure mechanisms is presented. Literature cites only one example where the Artificial Neural Network methodology was used, for relating water distribution damage

to natural hazards. This paper attempted to relate cold temperature hazards (air temperature, snow precipitation, and degree-days) to historical pipe damage (Bates et al., 1996). Unfortunately, neither was a discussion given as to the reason for including the specific factors, nor were quantitative results presented which demonstrated the accuracy of the model. The following parameters are presented with reasons supporting their inclusion.

Pipe Diameter

Review of pipe break literature indicates a strong correlation between the number of pipe breaks and the diameter of the pipe. A study of pipe breaks conducted in Winnipeg, Manitoba concluded that "the decreasing trend in pipe failure rate for cast iron pipe with increasing diameter is directly attributable to the increasing wall thickness and joint reliability with increase in pipe diameter. Larger wall thickness gives the pipe better structural integrity and improved resistance to corrosion failures" (Kettler and Goulter, 1985). Many other studies have also shown that a larger proportion of failures have occurred in the smaller diameter pipes [(Kitaura and Miyajima, 1996), (Bahmanyar and Edil, 1983) and (Rajani et al., 1996)].

Literature suggests that the pipe size also affects the mode of failure (O'Day, 1982). Smaller diameter mains (150 to 200 mm) often experience beam (flexural) failure because of poor bending conditions, however crushing failures (often longitudinal failures) are unlikely to occur due to the relative length-to-diameter ratio. Conversely,

larger mains (250 mm or greater) are likely to experience crushing failure, but are not likely to experience beam failure (O'Day, 1982).

Age of Pipe

The significance of pipe age as a determinant of pipe breakage is debatable. Some experts believe in a natural progression of occurrences of pipe breaks with age [(Goulter et al., 1990) and (Bates et al., 1996)]. Others have indicated that "studies show that age is not the major determinant of water main break rates" (O'Day, 1982).

Pipe Joint Type

Joint type is an issue since the type of joint will influence the susceptibility of the pipe to specific failures. A large part of this may be owing to the amount of flexibility and lateral constraint the joint provides, as well as the pipe joint's actual strength and its ability to resist corrosion. For cast iron pipes in the previously mentioned Winnipeg study, joint failure is predominant with bolted and universal joints (Goulter and Kazemi, 1989). Kitaura reported "joint separations for cast iron pipe occurred in the older lead and mechanical joints" (Kitaura and Miyajima, 1996). Morris Jr. speculates that certain types of bolted or welded joint connections are more susceptible to corrosion (Morris Jr., 1967).

Joint separations are leak failures where pipe joints become separated. The study of the failure of a 108-inch pipe demonstrated that "total pipe separations occurred because of large unrelieved thermal stresses and stress amplification cause by the eccentricity of the welded bell-and-spigot joints (Moncarz et al., 1987). A study by Milligan showed that "different filler materials will have differing abilities to accommodate a greater deviation off line before damage occurs" (Milligan, 1995).

Internal Pipe Water Temperature

Some literature speculates that a high differential temperature between the internal and external pipe wall can produce high temperature gradients. Under such conditions the inner and outer fibers will be subjected to different temperature drops, resulting in differential strains and circumferential stresses. The increase in hoop stress increases the likelihood of longitudinal failures (Habibian, 1994). A co-author of a study performed in Madison, Wisconsin did not agree that this mechanism was a problem for seasonally cold regions (Bahmanyar and Edil, 1983).

Operating Pressures

For circumstances where water pipeline pressure surges result in blowout (longitudinal) failures, a parameter that depicts changes in operating pressure (and therefore changes in circumferential stresses) is necessary. Such events may occur as pressure surges (during pump shut down or other normal pump operations) or during unforeseen events

(accidental valve closure, etc.). Such events may cause a water hammer effect, which is a likely explanation for a limited number of pipe failures.

Environmental factors exist which may intensify or conversely buffer the pressure experienced by the pipe. A study by Burrows and Qiu (1995) indicated that the presence of air pockets in pipelines can exacerbate surge peak. A typical example of this possibility is provided in Figure 3. Conversely, work by Rajani et. al. (1996) indicates that at lower ground temperatures, the elastic moduli of soils can increase significantly such that frozen soils will have a positive counteracting effect on the development of hoop stresses.

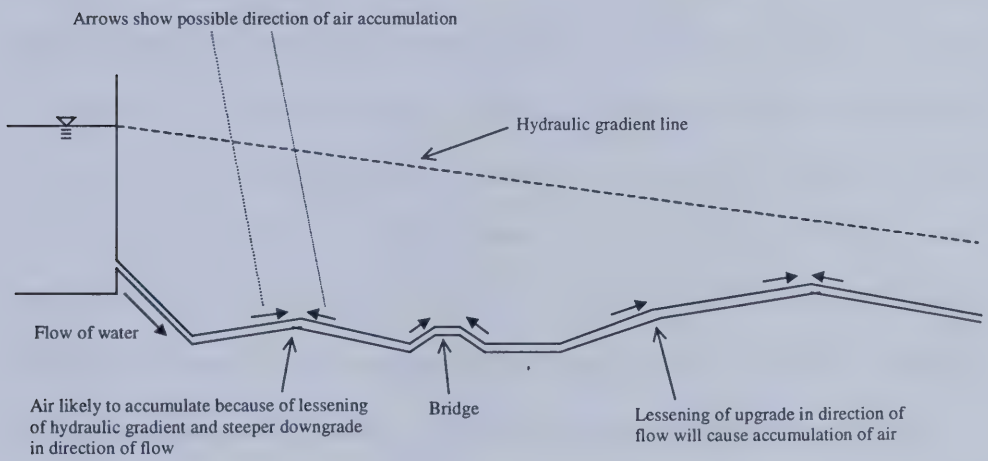


Figure 3. Effects of air pockets on operating pressures.
Adapted from (Burrows and Qiu, 1995)

Soil Type

The significance of the type of soil cannot be overlooked, as it is one of the most important factors, having effects on almost all of the above mechanisms. Its effects on frost heave, strength of soil-pipeline interaction strength, and external corrosion can be important for many failure mechanisms.

Frost susceptibility is defined as the rate at which frost penetrates the ground. It is generally regarded as one of the most important factors in characterizing frost heave action. Frost susceptibility is ranked greatest to least for soil types in the following order: silt, clay, sand, and then gravel. However, methods of further quantifying and thoroughly characterizing soils in terms of frost susceptibility are not consistent. Use of frost heave rate (mm/day), total frost heave (mm), frost heave ratio (ratio of frost heave rate to total frost heave) and segregation potential (to depict frost susceptibility (Kujala, 1993)) have been suggested. However, these types of measures are often difficult to find, or do not translate accurately from laboratory to field values (Konrad and Nixon, 1994). Others disagree, indicating that "for a constant pressure (load on the soil) the rate of heaving is independent of the rate of freezing…[which is] completely valid only for relatively permeable soil" (Penner, 1972). Therefore characterization of frost susceptibility, and hence frost heaving is difficult using field measurements.

The type of soil the pipe is located in is also important for the aspect of differential heaving and thaw settlement. If a pipe is located at the interface of two different soil

types, it has been shown that each soil will experience an uneven amount of frost heaving, and therefore have an influence on the amount of strain experienced by the pipe (Nixon, 1994). In the same manner, thaw settlement will lead to differential stress distributions on the pipeline.

Several studies have demonstrated that freeze-thaw cycles have effects on the mass transfer and physical properties of the soil [(Kurilko et al., 1989); (Kujala and Laurinen, 1989); (Rajani, 1992), and; (Pawluk, 1988)]. Results have also shown that textural changes caused by freezing and thawing of clayey materials may alter the mass transfer characteristics by two or three orders of magnitude, and as a result, affecting frost heave rates. However, the influence of cyclic freezing and thawing on sands is not observed at all (Kurilko et al., 1989).

Use of soil type to represent various soil properties is a difficult task due to the seasonal variance of several properties (deemed important to frost action, soil-pipeline interaction, and corrosion). In addition, many of the soil parameters influencing frost heave and corrosion will not be available due to impracticalities in monitoring. Therefore, these parameters may have to be inferred from soil type. Important parameters which are assumed to be constant with soil type are: soil thermal conductivity (Konrad and Morgenstern, 1980); Poisson's ratio (Shen and Ladanyi, 1991) and (Selig, 1988); and hydraulic conductivity (Anderson et al., 1984).

Soil corrosivity is a soil characteristic that must be considered for external corrosion predictions. Physical characteristics (particle size, friability, uniformity, organic content, color) have reflected corrosivity, based on observations and testing. Color has also been linked to corrosivity. Soil uniformity is important because of the possible development of localized corrosion cells. Corrosion cells may be caused by a difference in potential between unlike soil types, with both soils being in contact with the pipe (Smith, 1968). If it can be assumed that for a particular soil classification the approximate uniformity coefficient can be estimated, then the possibility of corrosion can be estimated.

Overburden Pressure

Overburden pressure is thought to be important due its ability to help characterize frost heaving and soil-pipeline resistance. It can be characterized by the depth of bury and soil density. To simplify assumptions, it will be assumed (for this study) that soil density is generally characterized by soil type.

With respect to frost heave action, overburden pressure is another important factor for ice lens formation (the others being: frost susceptible soil, freezing temperature and a source of water). Literature indicates that the overburden pressure is important for the rate of heaving [(Anderson et al., 1984); (Nixon, 1994); (Hu and Selvadurai, 1995), and; (Roy et al., 1992)].

Bury depth is an important factor for other reasons. From the perspective of soil-pipeline interaction, it has been demonstrated that the frictional soil resistance is affected by pipe diameter and bury depth (Rajani et al., 1995). Also, from the perspective of mode of failure, larger pipes are more susceptible than smaller pipes to crushing failure. This is due to bury depth, or the external loadings the pipe is subjected to (i.e. roadways, large structures (O'Day, 1982)).

Segregation Potential

Konrad thoroughly investigated segregation potential for characterizing frost susceptibility [(Konrad and Morgenstern, 1981); (Konrad, 1987); (Konrad, 1994), and; (Konrad and Nixon, 1994)]. This parameter is determined using laboratory measurements and is a proportionality constant comprised of the measured hydraulic conductivity (as pore-water velocity) and temperature gradient (Konrad, 1994). The value obtained depends on the stress and thermal histories of the soil deposit. This parameter may be especially useful as it combines two of the parameters that are important for frost susceptibility characterization, namely hydraulic conductivity (which also relates to water content) and temperature gradient (which relates to the freezing rate).

Soil-Pipeline Reaction Modulus (k_s)

The axial soil-pipe reaction modulus, k_s, is a parameter which describes the interactive resistive strength created by the soil and pipeline interface (k_s is typically expressed as MPa/m). Determination of this reaction modulus is done in one of several ways, either from elastic properties or empirical relationships from sand and clay (Rajani et al., 1996). These relationships are illustrated in the following equations:

$$k_s = \frac{G_s}{4(1 - \nu_s)\frac{D}{2}} \quad [1]$$

Where:

D is the external diameter of the pipe, mm

G_s is the soil shear modulus, MPa (N/mm^2)

ν_s is the soil Poisson's ratio

$$k_s = \frac{\alpha s_u}{u_y} \quad [2]$$

Where:

α is the adhesion coefficient

s_u is the undrained shear strength of clays, MPa

u_y is the displacement required to develop ultimate axial resistance, mm

$$k_s = \frac{0.5(\gamma H)(1 + K_0 \tan \delta)}{u_y} \quad [3]$$

Where:

γ_s is the submerged unit weight of soil, kN/m^3

H is the burial depth of water mains from surface to the centre line of the pipe, m

K_o is the coefficient of active resistance at rest

δ is the frictional angle between the pipe material and surrounding backfill

This soil-pipe resistance parameter is not a widely accepted parameter since investigation using this parameter is relatively new. Much of the recent research carried out has been performed by Biggar and Sego (Rajani et al., 1996). While this measure is ideal because it accurately measures soil-pipeline interaction, these values are not readily available. However, it is demonstrated that in general soil-pipe resistance increases with pipe roughness (Rajani et al., 1996).

Soil Elastic Modulus E_s

Literature has shown that the soil elastic modulus is a representative stiffness property for soil-pipeline interaction [(Rajani et al., 1996) and (Selig, 1988)]. Therefore, the soil elastic modulus will give a measure of the strength of the soil-pipeline interactions. This

is especially important if non-uniform soils are present. This parameter also may have a counteracting effect to the development of hoop stresses during water pressure surges.

Soil pH (and Soil Resistivity)

In order to characterize external corrosion, it is necessary to find parameters which indicate the corrosivity of the soil. Soil pH is a good general indicator of external corrosion since certain pH ranges allow for different corrosion mechanisms to occur. It has also been found that resistivity is a function of pH [(Morris Jr., 1967); (Booth et al., 1967), and; (Jarvis and Hedges, 1994)]. For that reason, only one of the two may be required for characterization.

Literature indicates a very poor correlation between soil type and soil resistivity (Dorn, 1989). For this reason, soil resistivity cannot be incorporated into the soil type parameter. This being the case, soil type and either soil resistivity or soil pH may be considered as potential input parameters, but inclusion of both soil resistivity and soil pH are not necessary.

Soil pH can be divided into three important ranges: 0 to 4, 6.5 to 7.5 and 8.5 to14 (Morris Jr., 1967). At a pH of 0 to 4, the soil acts as an electrolyte. In the neutral range, pH is optimum for sulfate reduction. At a pH of 8.5 to 14, soils are generally high in dissolved solids, and thereby yield a low resistivity [(Morris Jr., 1967) and (Jarvis and Hedges, 1994)].

Interpretation of soil resistivity field measurements is extremely important. Only when reading the resistivity in soil at the specific pipe depth can the interpretation (of corrosion potential) be made accurately (Smith, 1968). As a result, the ground water content and soil temperature must also be ascertained. With cast iron pipe, corrosion resistance is enhanced if there are dry periods during the year. This seems to permit hardening or toughening of the corrosion scale or products, which then become impervious and serve as a better insulator (Smith, 1968). Also demonstrated by Smith was that resistivity will vary with the soil temperature. As the soil approaches freezing, resistivity will increase greatly, and thus a reliable reading may not be possible. If resistivity is to be measured, consideration must be given to a lack of consistent readings between field and laboratory measurements. Therefore it is necessary to assign ranges of resistivity, rather than specific numbers (Smith, 1968). Common ranges for soil resistivity are given in Table 1.

Table 1. Common ranges of resistivity for soils.
Adapted from (Dorn, 1989)

CORROSION CLASS	RESISTIVITY (METRE-OHMS)
Severly Corrosive	0 to 5
Very Corrosive	5 to 10
Corrosive	10 to 30
Moderately Corrosive	30 to 100
Slightly Corrosive	Above 100

Soil Aeration (Redox Potential)

This parameter is useful in characterizing the potential for bacterial corrosion. It is an established fact that sulfate-reducing bacteria can live only under anaerobic conditions (a redox potential greater than100 mV indicates sufficient soil aeration so as not to support sulfate producers (Morris Jr., 1967). However, Jarvis and Hedges (1994) contradict the above, stating that values less than 400 to 430 mV indicates a suitable environment for sulfate-reducing bacteria. Overall, while there may not be a general agreement on the range for which sulfate corrosion is favorable, there is agreement that this measurement provides potential for characterizing bacterial corrosion potential.

Soil Water Content

Use of the soil water content parameter is important from several aspects. As mentioned earlier, the rate of frost heave is controlled by the availability of free water (McGaw, 1972). It is also important for external corrosion.

From the perspective of frost heave, it has been stated that the availability of a water source is one of the necessary elements required for ice lens growth. In the absence of a nearby ground water table, focus then shifts to the availability of water present in the soil itself, i.e., soil water content. In reality, the water content may be a possible surrogate measure for water table depth, as water may enter the soil above by capillary suction.

From the perspective of external corrosion, soil corrosion aggressiveness has been related to moisture content. Soils with a moisture content above 20 percent (wet basis) are thought to be particularly corrosive (Jarvis and Hedges, 1994). Another study cited a correlation between soil aggressiveness and an optimum water content at a minimum resistivity. Therefore, these results substantiate moisture content as a measure of soil aggressiveness (Booth et al., 1967).

Anderson and Tice (1972) provided a study that demonstrated that water content may be related to soil temperature and specific surface area. In this study, an empirical equation was devised relating the unfrozen water content of partially frozen soils to the soil temperature and specific surface area. Results comparing computed water contents and experimentally obtained values showed good agreement, particularly for temperatures below –5 ^{o}C. Another study demonstrated how water content is also affected by the salinity of the soil (Jones, 1995). These observations again demonstrate the importance of soil type as a general indicative parameter.

Cluster Indices

As mentioned earlier, several studies have been performed that focus on the analysis of causes of cast iron water main failures. There has been particular focus on the temporal and spatial correlation of break events. The Winnipeg study performed by Goulter indicates that pipe failures often occur in clusters, located in relatively close proximity of other breaks [(Goulter et al., 1990); (Goulter and Kazemi, 1988), and; (Goulter and

Kazemi, 1989)]. The same study shows that the likelihood of a break occurring decreases with time from the break of another pipe in the area [(Goulter and Kazemi, 1988) and (Goulter and Kazemi, 1989)]. Further investigation of these phenomena is needed, to determine its significance.

Air Temperature

Air temperature is an important parameter since it characterizes the change in climate, and on a smaller scale, the change in seasons. It is integral for characterizing the potential for frost heaving and soil-pipeline interaction stress generation. It also affects the measurement and stability of soil parameters.

Deterministic models show there is a relationship between air and ground temperatures (see Figure 4). However, heat transfer from ground to air requires time due to the reduced thermal conductivity of the soil (compared to air). Since time is required for the ground temperature to equilibrate to the ground surface temperature (i.e. temperature at the ground-air interface), ground temperature can be considered a function of depth. In this manner, air temperature indirectly affects pipe failure mechanisms in two ways:

1. temperature-induced contraction, and;
2. frost heave mechanics (O'Day, 1982).

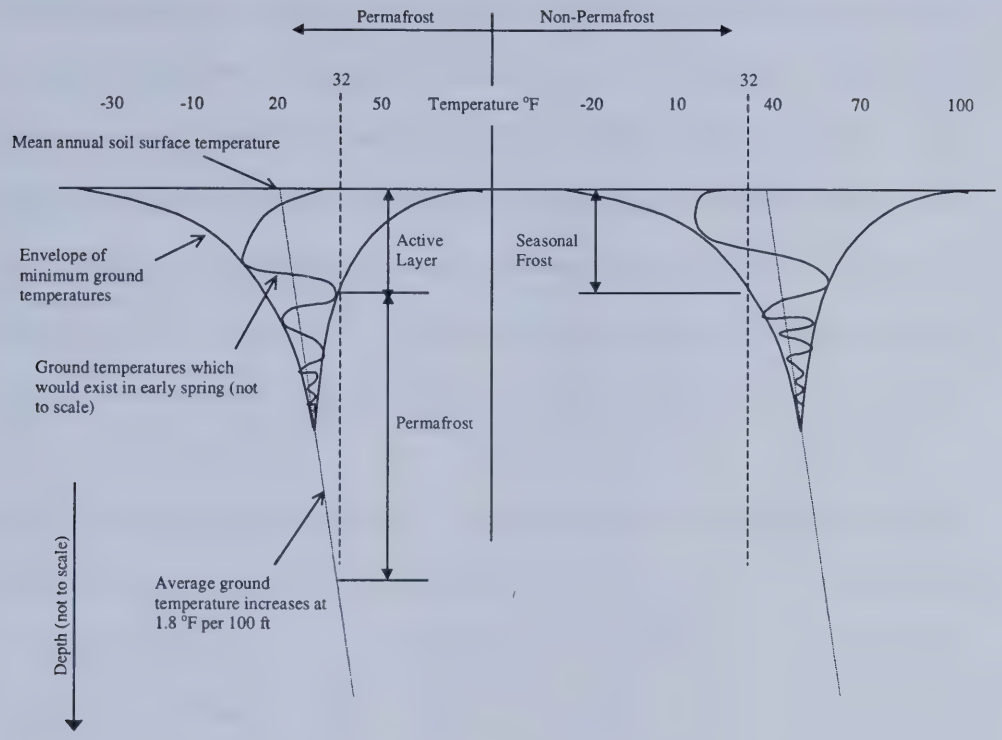


Figure 4. Whiplash curve for ground-to-air heat transfer. Adapted from (Phukan, 1985)

As illustrated in the above deterministic model, air temperature is not indicative of the ground temperature unless the ground temperature is considered as a function of time and depth of interest. In using the Artificial Neural Network modeling approach, presenting the model with time-series data may allow the model to characterize the air-to-ground temperature transition effects without using the actual ground temperature (at pipe depth). This of course assumes that the conductivity of different soils is accounted for.

It is speculated that the daily drop in air temperature will be indicative of the rate of frost penetration [(McGaw, 1972); (Miller, 1972); (Anderson and Tice, 1972); (Penner, 1972); (Anderson et al., 1984); (Shen and Ladanyi, 1991); (Hu and Selvadurai, 1995); (Bahmanyar and Edil, 1983), and; (Roy et al., 1992)]. Some of these aforementioned authors have also reported that there is a correlation between a drop in air temperature and an increase in pipe breaks. One conflicting opinion, expressing that air temperature drops were not responsible for pipe breaks, was found (Habibian, 1994).

Use of air temperature (or differences in air temperature) may also be considered as a surrogate measure for freezing and thawing indices (Boyd, 1973). Use of these indices in prediction models, and use of air temperatures in time series may give a more realistic representation of climatic changes.

Consideration of air temperature is also important for its effect on many soil properties. Constant monitoring of changes of soil water content, hydraulic conductivity, undrained strength, elastic modulus (Shen and Ladanyi, 1991), resistivity, and depth of consolidation is not feasible. It is anticipated that use of the air temperature parameter in conjunction with precipitation and soil type will allow the ANN model to account for these changes in parameters.

Precipitation (snow and/or rain)

Snow is indicative of the insulating effect on ground temperature, as the snow will allow for the entrapment of heat into the ground. Rain precipitation coupled with the soil type may be indicative of moisture content or hydraulic conductivity if these parameters are not measured regularly. Some literature indicates that corrosion resistance is enhanced during dry periods of the year (Smith, 1968). Rain precipitation may also indicate an abundance of water supply [(Penner, 1972); (Anderson et al., 1984); and; (Roy et al., 1992)]. Therefore, inclusion of this parameter may be necessary to help characterize climatic changes as well as to infer adjustments to soil parameters.

Rain precipitation may be a significant factor if water main breaks can be related to the swelling and consequently the instability of saturated clay soils during heavy rainfall events. This is based on the fact that many clay soils swell and shrink according to the soil moisture content to a high degree and exhibit a high plasticity and cohesion (Clark, 1971).

Summary of Parameters

Based on the literature reviewed, a conceptual list of causal and influencing parameters is presented in Table 2. These are input parameters which should ideally be included in the ANN model. Analysis of data availability or reliability will determine the feasibility of their inclusion.

Table 2. Summary of Influencing and Causal Factors.

FACTOR	POSSIBLE INFLUENCES ON PIPE BREAKS
Pipe diameter	Pipe strength characteristic
Age of pipe	Possible corrosion factor
Pipe joint age	Strength and rigidity of connections
Internal pipe water temperature	Pipe-wall temperature differential stresses
Operating pressures	Water hammer; high water pressures
Soil type	Frost action; Soil-pipeline interaction; Corrosion related
Overburden pressure	Frost heave characterization; Soil-pipeline strength
Segregation potential	Frost susceptibility
Reaction modulus	Soil-pipeline resistance
Soil elastic modulus	Soil-pipeline interaction
Soil pH (or resistivity)	Corrosion parameter
Soil aeration (redox potential)	Corrosion potential
Soil water content	Frost action mechanics; corrosion
Cluster indices	Pipe break phenomena
Air temperature	Frost action; Soil-pipeline interactions; Soil parameter characterization
Precipitation	Soils stability (especially clays); ground temperature insulating effects

2.4 Artificial Neural Networks Overview

With the influencing and causal factors now outlined, it is demonstrated that there are wide arrays of factors that have potentially significant effects on each type of pipe failures. There are established relationships and interactions between many of these factors. It can also be mentioned that although many influencing factors are identified, there is uncertainty as to their roles for causing cast iron pipe breaks. Expert knowledge as to causes of failures is reasonably advanced yet past attempts at modeling these failures have not yielded satisfactory conclusions. An alternative method of modeling, using Artificial Neural Networks is presented which has the potential to model the pipe breaks.

2.4.1 Background Information and Application

The Artificial Neural Network modeling technique, though not conceptually new, has only recently been explored in many fields of civil and environmental engineering because of its requirements for intensive computing capabilities. With the advancements in computing technology, this modeling technique is gaining popularity for its abilities to deal with problems having non-linear solutions, and particularly with its ability to forecast events.

This modeling approach is generally regarded as one of the best at extracting concepts from historical data, and has a strong ability to learn, and thus has the ability to forecast

future events within its study domain. Since the ANN modeling technique utilizes an organizational structure that has numerous interconnections, it does not rely upon deterministic mathematical equations. As a result, its structure is configured such that it can handle complex, non-linear problems.

Appropriate application of Artificial Neural Networks requires that the following characteristics of the problem application exist:

1. The algorithm required to solve the problem is unknown or expensive to discover;
2. Heuristics or rules required to solve the problem are unknown or difficult to establish and;
3. The application is data intensive and a variety of data are available (Zhang, 1996).

As previously mentioned, earlier attempts at developing an analytical procedure using conventional methods for predicting water main failure have not proven successful. The possibility of modeling pipe break failures using ANN exists provided that adequate, easily accessible information can be collected. Ideally, a comprehensive model will involve data inputs from the causal and influencing parameters described above.

2.4.2 ANN Model Development Process

This general process of developing an Artificial Neural Network is composed of four interdependent stages:

1. Source Data Analysis;
2. System Priming;
3. System Fine-Tuning; and
4. Model Evaluation.

Source data analysis involves identifying and preparing the potential input parameters for use in the Artificial Neural Network models. Description of the considerations for this model study is to be detailed in Section 3.1. System priming involves determining which of the potential input parameters will be most appropriate for the problem study. The system fine-tuning involves adjusting the Artificial Neural Network model structures to optimize learning of the input data presented. The system priming and system fine-tuning stages are often done concurrently and are described in Section 3.2. In the final stage, Model Evaluation, the models are evaluated both qualitatively and quantitatively.

These stages will be described in more detail in Section 3. For the ANN modeling portion, the program *Neuroshell 2*, developed by Ward Systems Group, Incorporated was used (Ward Systems Group, 1993).

2.4.3 ANN Model Structure

General Concept

Artificial Neural Networks is one classification of Artificial Intelligence (AI). It is a "black box" methodology that typically "learns" by comparing an input pattern or sequences of input patterns, to the outputs. Relationships between model input and model output are formed using an error correction algorithm, which attempts to minimize the error between model prediction and the actual events. Error corrections are distributed amongst the hidden neurons within the model. When presented with a wide range of appropriate input patterns and a suitable model structure, these models are able to intrinsically learn the underlying logic features and importance of these patterns.

ANN technique is a form of artificial intelligence that simulates what we think we know about how the brain works (Schmuller, 1990). An Artificial Neural Network is composed of a set of simple processing units called "neurons" (as illustrated in Figure 5), each capable of only a few computations such as summation and threshold logic (Garrett et al., 1992). These neurons are arranged in layers, with the neurons from each layer being interconnected to one another. This configuration lends itself to self-organization and learning.

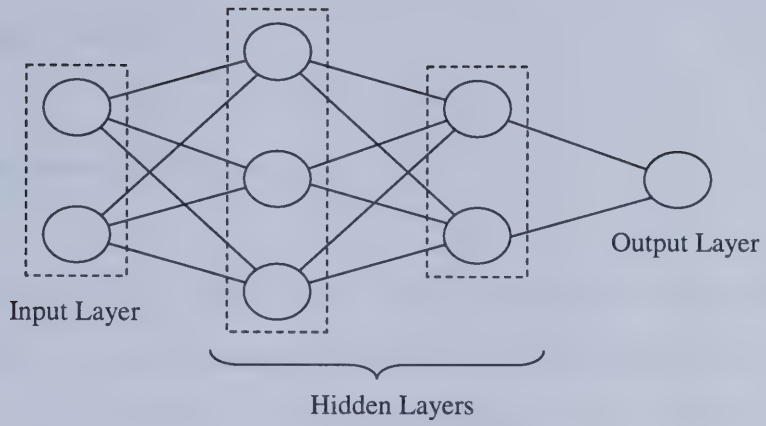


Figure 5. Typical ANN backpropagation structure.
Modified from (Sacluti et al., 1998)

The Artificial Neural Network learning process involves the entry of significant input parameters into the model, with the output parameters known (for "supervised" learning). Conversely, "unsupervised" learning is where the output parameter is unknown. Input data (often modified and/or placed in time series) enters each individual neuron with an initial weight value. Depending on the scaling function (or activation function in subsequent layers), a new significance value is assigned to the output signal. These values are conveyed to other interconnected neurons in subsequent layers until an output parameter is determined. The actual learning occurs when feedback iterations are performed, and the model begins to organize itself according to the data it is presented.

Optimal and organized model development involves selecting appropriate input parameters and suitable ANN model features. These structures may be internal or external to the neuron units. Those structures internal to the neuron are the scaling and

activation functions. The most significant model feature external to the neuron is the type of learning architecture.

Scaling and Activation Functions

Data initially fed into the model must be scaled from their numeric range into a relative range for which the network can deal with more efficiently (Ward Systems Group, 1993). The model structure performing this is called the "scaling function". Scaling may be linear (Equation 6), but may also use the non-linear scaling functions logistic (Equation 4) or tanh (Equation 5). These non-linear scaling functions will tend to group data at both the higher and lower limits of the original data range. *Neuroshell 2* sets the default scaling function to the linear function.

Activation functions are structures in layers subsequent to the input layer (i.e. hidden and output layers). They dictate how the individual neurons pass neuron output weight values from the summed neuron input weight values of the previous layer. The activation function maps the inputted sum into the output weight value, which is then passed onto the succeeding layer.

NeuroShell 2 provides a number of activation functions that allow for flexible application to problems. The more significant activation functions are: logistic (Figure 6), linear (Figure 8), tanh (Figure 7), Gaussian (Figure 9), and Gaussian complement (Figure 10).

The default setting for models is the logistic function. This function has been found to work best for most neural network applications (Ward Systems Group, 1993), however, there are always exceptions to this.

The logistic function is mathematically described as:

$$f(x) = \frac{1}{(1 + \exp(-x))} \qquad [4]$$

This is graphically represented by:

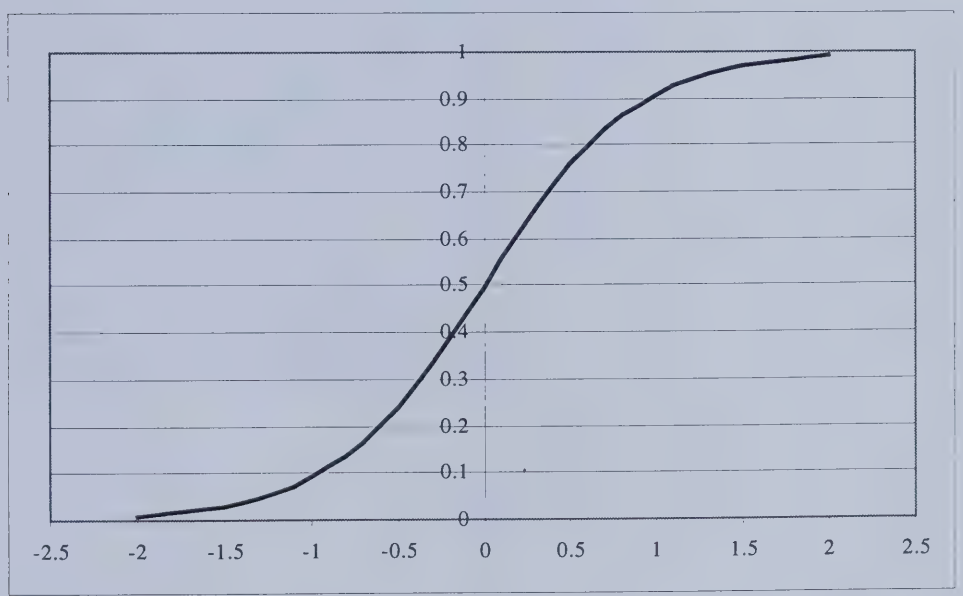


Figure 6. Logistic activation function.
Adapted from (Ward Systems Group, 1993)

The hyperbolic tan function is:

$$f(x) = \tanh(x) \qquad [5]$$

and is illustrated by:

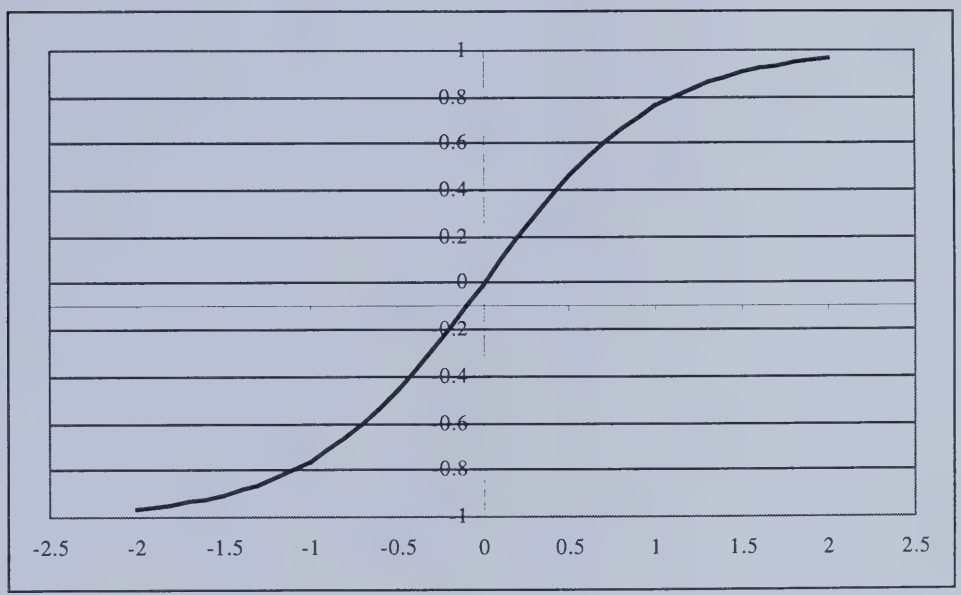


Figure 7. Tanh activation function.
Adapted from (Ward Systems Group, 1993)

The linear activation function is given by:

$$f(x) = x \qquad [6]$$

and is graphically represented by:

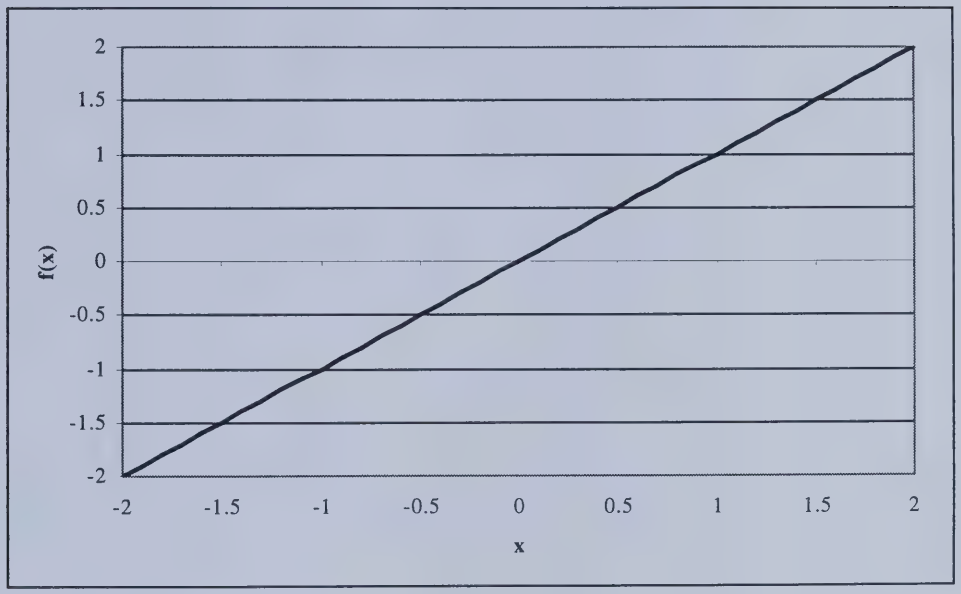


Figure 8. Linear activation function
Adapted from (Ward Systems Group, 1993)

The Gaussian activation function is mathematically described by:

$$\text{Gaussian} = \exp\left(-x^2\right) \qquad [7]$$

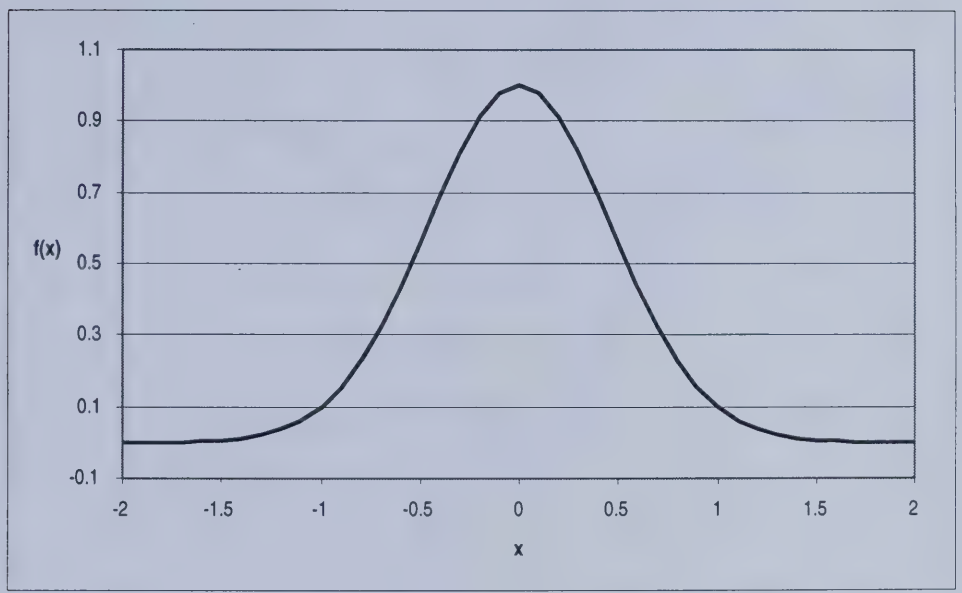


Figure 9. Gaussian activation function
Adapted from (Ward Systems Group, 1993)

The Gaussian Complement activation function is described by:

$$\text{Gaussian complement} = 1 - \exp\left(-x^2\right) \qquad [8]$$

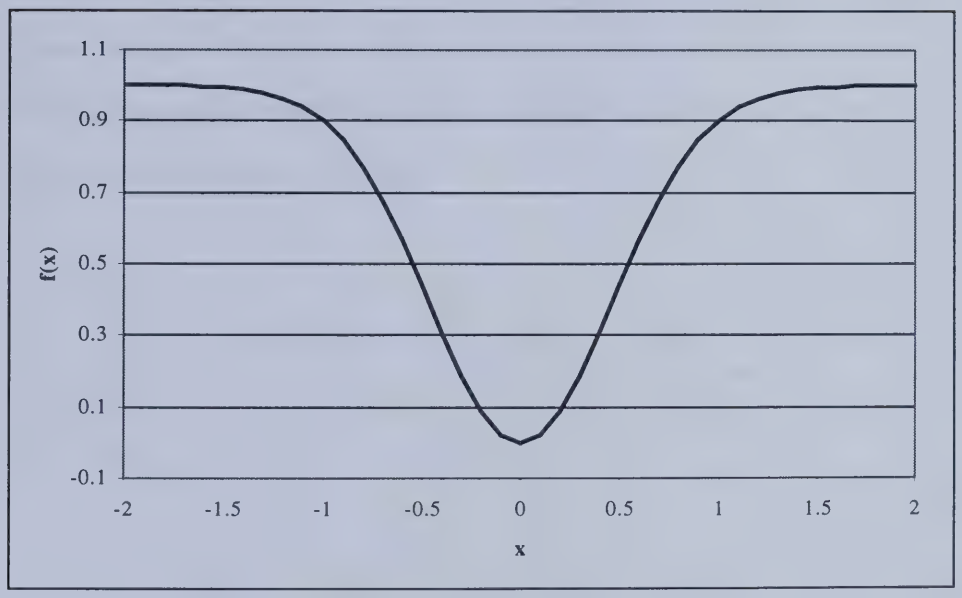


Figure 10. Gaussian complement activation function
Adapted from (Ward Systems Group, 1993)

Learning Architectures

As stated earlier, the program used for model development, *Neuroshell 2*, includes several different types of Artificial Neural Network supervised learning architectures. These architectures include the standard Backpropagation networks (Figure 11), but also includes Probabilistic Neural Networks (PNN) and General Regression Neural Networks

(GRNN). Within each of these learning architectures, especially the Backpropagation networks, unique features of the network structure give way to distinct model types such as Jump Connections networks (Figure 12), Recurrent networks (Figure 13), and Ward networks (Figure 14). This subset of backpropagation networks is designed to provide flexibility in design of the network, and vary the method of data presentation (Ward Systems Group, 1993). This flexibility may allow the network to capture specific features of the data set or problem study, not as easily captured with a standard connection backpropagation network.

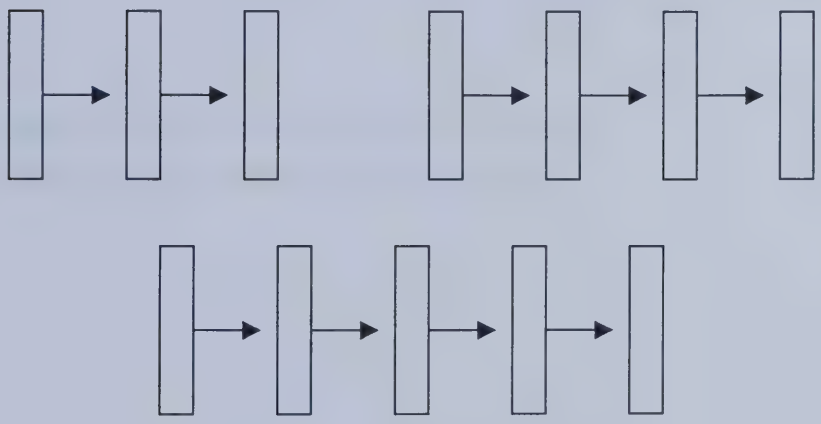


Figure 11. Standard connection network structure.
Modified from (Ward Systems Group, 1993)

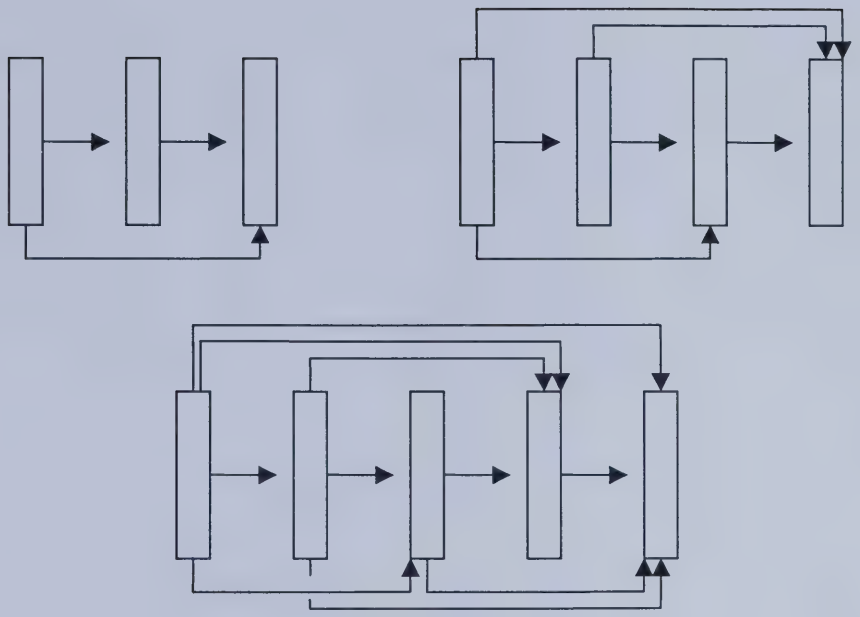


Figure 12. Jump connection network structures.
Modified from (Ward Systems Group, 1993)

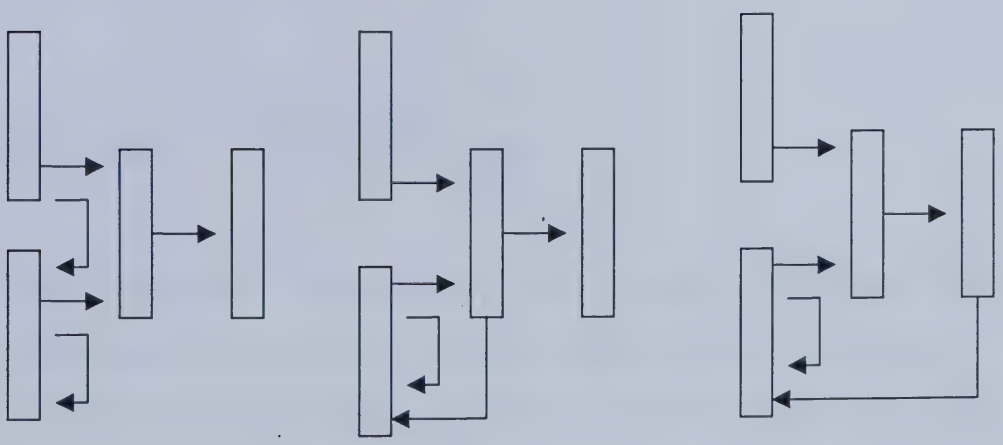


Figure 13. Recurrent network structures.
Modified from (Ward Systems Group, 1993)

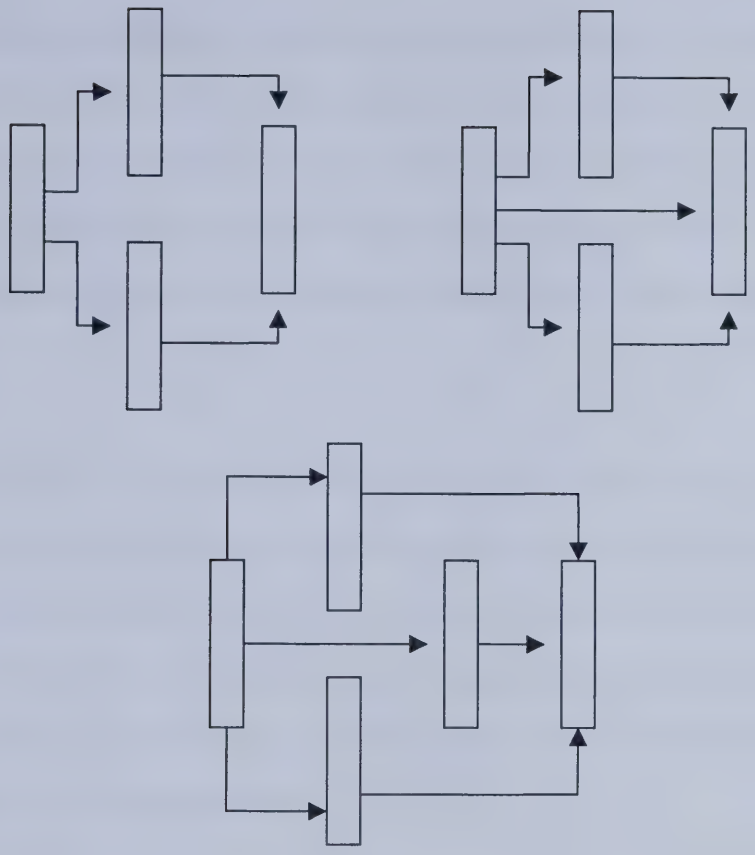


Figure 14. Ward network structures.
Modified from (Ward Systems Group, 1993)

The Backpropagation architecture offers the greatest amount of flexibility in *NeuroShell 2* since there are a number of network options. Standard Connection networks are the simplest form of ANN model, with each layer of neurons connected to the adjacent layer. Jump Connection networks allow for more involved linking, such that every layer is connected to each other layer, not solely to the adjacent layer. The Recurrent Networks have an additional input layer which stores the contents of the previous pattern that was

trained, allowing the network to see previous knowledge it had about previous inputs (Ward Systems Group, 1993). A final backpropagation type model offered in *NeuroShell 2* is the Ward Network, developed by the supplier of the program. This network employs the use of varying activation functions, designed to identify different features of the data set. Thus, the output layer receives different "views of the data" (Ward Systems Group, 1993).

The Probabilistic Neural Networks and General Regression Neural Networks are both three layer network architectures in which the input patterns are presented to the input layer, and each individual input pattern is retained by one or more hidden layer neurons. The output from the individual pattern is either categorized according to a probability density function (for PNN) or presents a continuous value output (for GRNN) based upon a comparison to all retained patterns.

The initial layer of all networks is comprised of the model's input parameters, with each input parameter assigned to an input neuron. These neurons are collectively called the input layer. The final layer is described as the output layer, and is composed of a single output (sometimes multiple outputs) that is (or are) being modeled. This is collectively called the output layer. The layers in between are described as "hidden layers" because, while each neuron serves as a unit-process decision (each neuron makes a decision based on the input it receives from other neurons) these decisions are intrinsic to the model, and these neurons' outputs are not readily seen. The combination of the hidden neuron

outputs and interconnections between neuron layers culminate in the decision-making process of the model.

2.4.4 ANN Model Learning Mechanics

The ANN models employed for this study used an error backpropagation algorithm for learning. For this type of learning, input patterns fed into the model produce an initial model output. The resulting output is then compared to the actual outcome (corresponding to the input pattern) to determine the model's prediction error. This error is then fed back into the hidden layer neuron interconnection weights. The error is batch corrected, and errors are distributed throughout the neuron connections. These modified weights attached to each incoming neuron signal, produce new output weights, which are governed by the type of activation functions utilized. These new output values are then transmitted to the neurons of the following interconnected layers, until a new output signal is derived at the output neuron. This error correction cycle continues until either the prediction reaches a minimum value (determined by a user-specified maximum iteration period) or the prediction error is below a user-specified error limit. Due to the complexity of the neuron interconnections, connection weights will continue to develop until a set of stable connection weights is found.

2.4.5 ANN Data Set Presentation

As stated earlier, ANN models learn by the presentation of input and actual output data patterns. When developing the model, it is important to realize that the purpose of presenting the model with various data patterns is to allow the model to intrinsically extract logic concepts from the data set. For this to occur, information presented must be representative of the full range of events, and a wide range of different patterns. The method by which the model is presented with the data is also important.

Three data sets are used: training, testing and production sets. The training data set is the set of patterns from which input and output patterns are initially entered into the model to train the ANN model. The developing model cyclically compares itself against a test data set, from which the model calculates its progress (where the goal is to minimize the error between the actual output, and the output of the model). The development and comparison cycle continues until there is a minimum specified error in prediction, or the model is unable to progress further. The model can then be tested against the production set for model verification. The production data set is a set of data points, which the trained model has never seen before. This data set is used for quantitative measurement of model learning ability and feature extraction capability. The pattern file contains all of the data pattern sets, including training, test and production sets.

3.0 METHODOLOGY

When assessing the feasibility of Artificial Neural Network methodology for research into pipe breaks, the modeling process must involve demonstration of knowledge and a logical experimental thought process. This section demonstrates these abilities through the collection, evaluation and use of data, and logical use of Artificial Neural Networks.

3.1 Data Collection

The effectiveness of an ANN model depends on the availability and reliability of the input data. Finding data that represents or corresponds to the possible factors reviewed was important for representing the physical cause-effect relationships. The reliability of the data is measured by the amount of "noise" inherent in the data. Noise are data patterns that contain inaccuracies and discrepancies, which does not allow the model to make proper associations between input and output patterns. Use of data with little apparent noise would result in a more accurate and precise model. As a result, precision in monitoring and collection of data was analyzed.

Data collection involved evaluating all available data based on accessibility, relative ease of obtaining long-term relevant data, and the prospect of future availability of the same type of data for future models. This data must have characteristics that are significant for model convergence. If all the proposed model input parameters are used for the model, the run times for model training will be exceedingly long, and hence would be an

inefficient use of time. Also, if insignificant (or inappropriate) data is not eliminated initially, the redundant input parameters will be treated as "noise" by the ANN model, and as such may decrease the likelihood of model convergence.

3.1.1 Parameter Collection and Analysis

Artificial Neural Network modeling requires not only representative data, but meaningful forms of the data. For this reason, manipulation of the raw data may be required to make input patterns more meaningful to the model. In cases where multiple failure mechanisms depend on a specific parameter, but for different physical representations, it becomes necessary to represent the same raw data in different forms to reflect the significance of these different representations. For this purpose, input patterns are changed to be more indicative of the underlying failure mechanism. A review of the plausible model input parameters and justification of their transformations are presented below.

Because of the nature of pipe failures, literature indicates that to fully predict these occurrences, it is necessary to have a wide range of representative data parameters. Due to limitations in the collection of input data, it became necessary to restrict the scope of the output being predicted. As mentioned earlier, the most important factors include pipe characteristics, pipeline operating characteristics, soils characteristics, soils properties, environmental properties and cluster indices.

Investigations into available raw data indicated that a large number of the suggested input parameters would not be available. Data for the various ideal input parameters that met the requirements (reliable and available in reasonable abundance) was difficult to collect.

Source data relating to individual pipe characteristics was found in water main leak or break repair reports. This report is standard documentation to be included with any water main repair performed. The information contained on this report is detailed, since it includes:

- pipe break location (street, avenue and distance from property lines);
- pipe depth at bury and apparent frost depth;
- report and repair dates;
- pipe material;
- size of pipe (pipe diameter);
- nature of break (i.e., longitudinal, transverse, etc.);
- apparent cause of break;
- condition of pipe and/or coating or wrapping and;
- site map of break.

This data was obtained from hardcopy documents stored at the water yard, as this level of detail is not typically entered in computer spreadsheet form.

Air temperature and precipitation (rain and snow) raw data was available from the closest weather monitoring stations. In this study, meteorological data from the municipal airport was used, since this is a city-central monitoring station, and located fairly close to the study area. Operating pressures were available, although they were only average daily pressures, and pump shutoff and water hammer events were not typically logged. Water temperature values were available for the supplying reservoirs and water treatment plants. Soils characteristics (soil type, soil parameters, corrosion-related parameters) were not available in sufficient detail or with enough frequency in collection to be useful.

The study was limited to modeling a city subdivision within the City of Edmonton's water distribution system (Figure 15). The selected area was the Calder subdivision (refer to map on Figure 16). This study area was a high-density break area for which data from 1972 to 1994 were collected and input into a computer spreadsheet. It was unknown why this particular area was experiencing a high break density when surrounding areas had similar pipe infrastructure and soils characteristics, yet were not experiencing similar break activity.

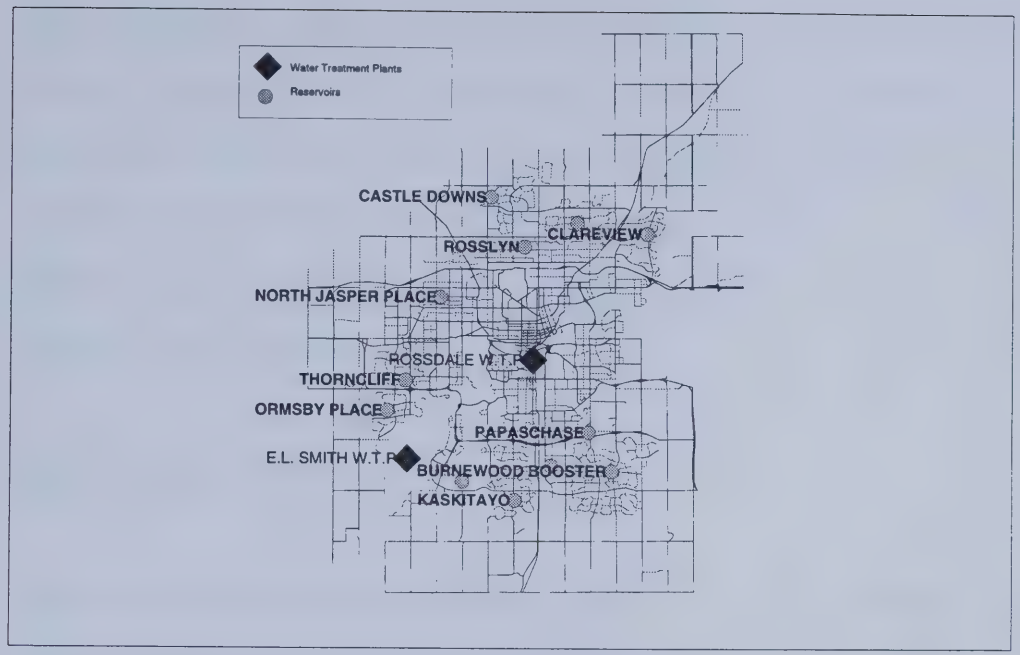


Figure 15. General Map of Edmonton Water Distribution System

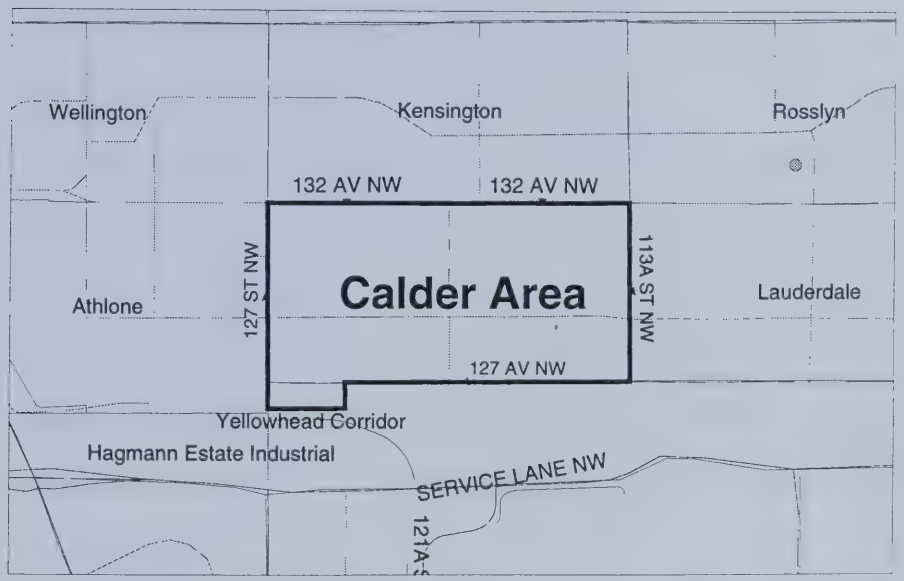


Figure 16. Calder Sub-division Study Area

Because information was either lacking or too general for application within the scope of this study, simplifying assumptions had to be made. For example, it was assumed that the subdivision was a uniform soil type. Soil maps of the Edmonton area (Kathol and McPherson, 1975) indicated that the Calder area was composed entirely of a Malmo Silty Clay Loam. These maps also indicated that the majority of the Edmonton area is composed of a silty clay loam.

Output Analysis and Format

Due to the unavailability of some potential raw input data (instantaneous operating pressures, many soils parameters), it was necessary to model only those pipe failures for which causal and influential factors could be easily obtained. The abundance of meteorological data, water temperature, and average operating pressures allowed for effective modeling of frost heave, soil-pipeline interaction, and pipe wall differential stress failure mechanisms. Because many failures were not directly attributed to corrosion, but were intuitively contributing causes to the above failures, they were included in the modeling. Failures as a result of indeterminate causes (i.e. clamp failures) were excluded from modeling. Those failures caused by unpredictable pressure surges (such as water hammer events) could not be excluded, since operational log data did not indicate when such events actually occurred. However, these events were typically infrequent.

A water main pipe failure is defined as an event where the leaking water pipe requires repair. Analysis of the raw break data (water main repair forms) was performed in order to determine if individual failure modes were seasonal. Since break information was available from 1972 to 1994, analysis was performed on all break data. Results illustrated that transverse failures (Figure 17) and diagonal failures (Figure 18) appeared to be seasonal, as the highest frequency of breaks occurred in the colder seasonal months. Longitudinal failures (Figure 19), blowout failures (Figure 20), and clamp failures (Figure 21) do not seem to have a seasonal pattern.

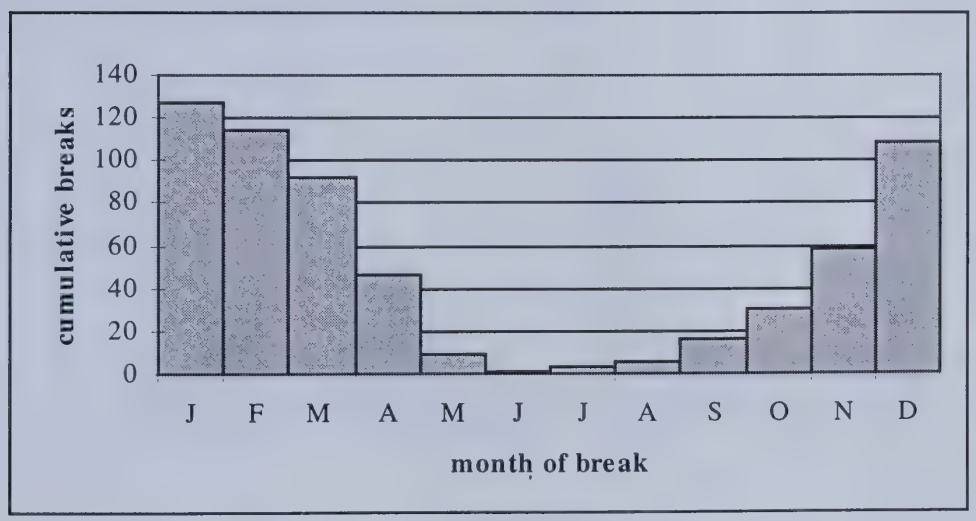


Figure 17. Monthly transverse failures, 1972-1994.

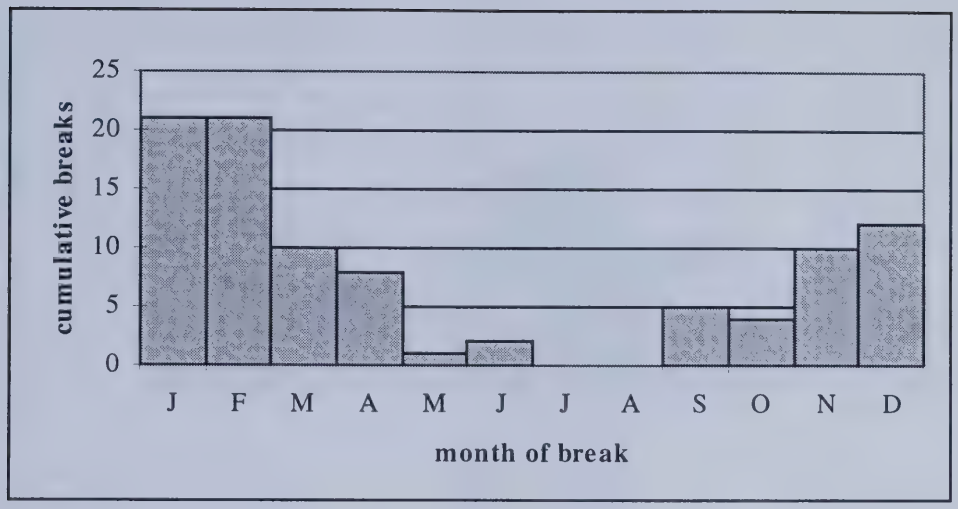


Figure 18. Monthly diagonal failures, 1972-1994.

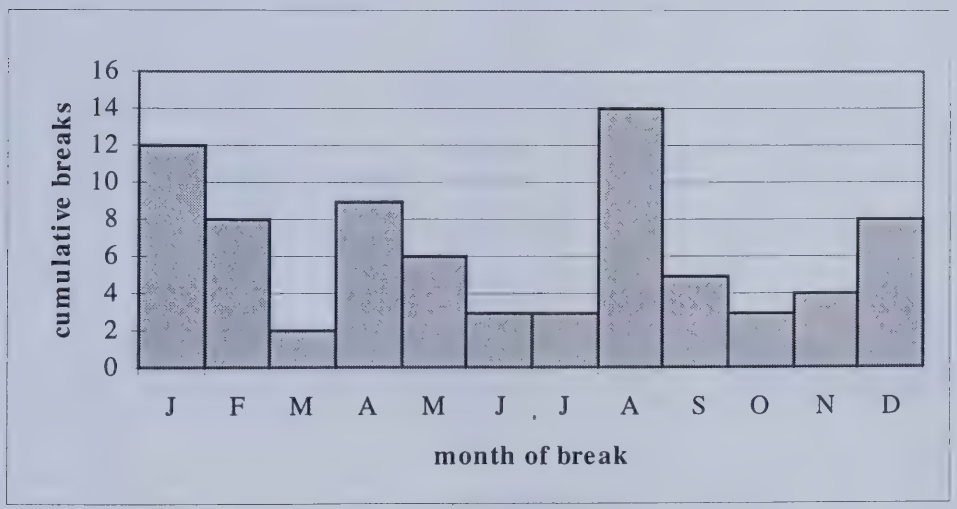


Figure 19. Monthly longitudinal failures, 1972-1994.

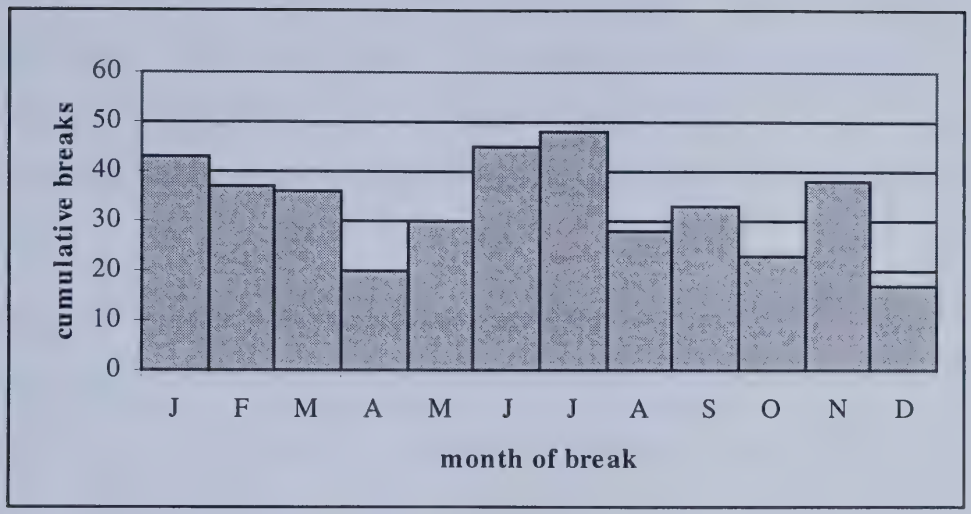


Figure 20. Monthly blowout failures, 1972-1994.

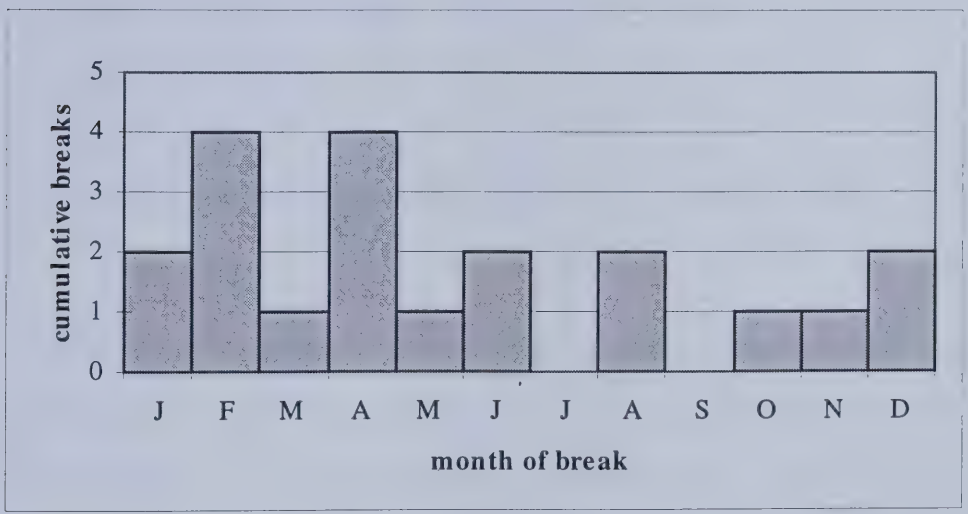


Figure 21. Monthly clamp failures, 1972-1994.

The number of 150 mm diameter pipe breaks was decided upon as the output parameter of primary focus since this pipe size accounts for 314 of the 564 pipe breaks (55.7

percent) on record for the period of 1985 to mid-1991. Table 3 lists the breakdown of pipe breaks according to pipe diameter. The reason behind limiting the scope of the study to a single pipe diameter was also to eliminate possible confusion in pipe failure mechanisms, since literature has described different failures modes for different pipe sizes.

Table 3. Pipe breaks sorted by diameter, 1985-1991.

Pipe Diameter	Number of breaks	Percentage of total
100	49	8.7
150	314	55.7
200	153	27.1
250	41	7.3
300	7	1.2
Total	564	100.0

It must be kept in mind that the model being developed will be able to predict the probability of pipe failures for a general area, but is not meant to predict the probability for individual pipes. As stated in the Section 1.2, the purpose of this model study is to demonstrate the possible utility of using ANN modeling for predicting pipe breaks. As such, further research and model development will likely be necessary.

Collection and Analysis of Raw Input Data

Following the determination of the availability of data was the determination of the reliability of the data. Also to follow was the transformation of the data into input forms that reflected the physical causes and influences on the different failure modes. Investigation revealed that the best potential for input parameters was available for air and water raw data, average operating pressures, precipitation data and temporal clustering indices. Analysis of the raw data, data manipulation and their justifications are given below.

Air Temperature

Use of this data as a substitute parameter for ground temperature was warranted due to the virtual absence of ground temperature monitoring. Ground temperature is an excellent indicator of the physical processes affecting pipe breaks. Aside from the aforementioned hoop stress conditions it creates, it is also an important factor for modeling frost heave (Anderson et al., 1984) and soil-pipe interactions (Yen et al., 1981). Thus, it was seen of paramount importance to characterize ground temperature data in order to characterize the different failure mechanisms. The most appropriate surrogate parameter was air temperature, given that this can be used to characterize ground temperatures in time series (to account for time lags in temperature reaching the lower ground depths).

Air temperature was thoroughly analyzed through graphical methods in order to determine a time-series correlation. Either 1-year or 2-year analysis periods were selected to maintain reasonable accuracy of the analysis. A graphical analysis of average daily temperatures (Figure 22) was performed. As expected, this analysis demonstrated that temperature had seasonal variations. A first and second difference was taken from the mean daily temperature (Figure 23 and Figure 24, respectively) in order to determine if a time lag relationship was apparent. This was evident, and it was decided that the model input parameter for air temperature must include a time lag component.

An analysis of the 7-day temperature change was also performed, and again, this analysis proved that a time lag component was necessary (Figure 25). This judgement was reasonable, since ground temperature transition models (e.g. Figure 4. Whiplash curve for ground-to-air heat transfer.) support this conclusion. Temperature transition models, such as this one, propose a slow thermodynamic transition of ground temperature from surface to increasing depth. Therefore, using a 7-day temperature lag was thought to be reasonable.

To further determine which representations of the raw data are most indicative of the physical process, the information will be used in the Artificial Neural Network model. The results of the ANN model show which form of the parameter is most effective for model input.

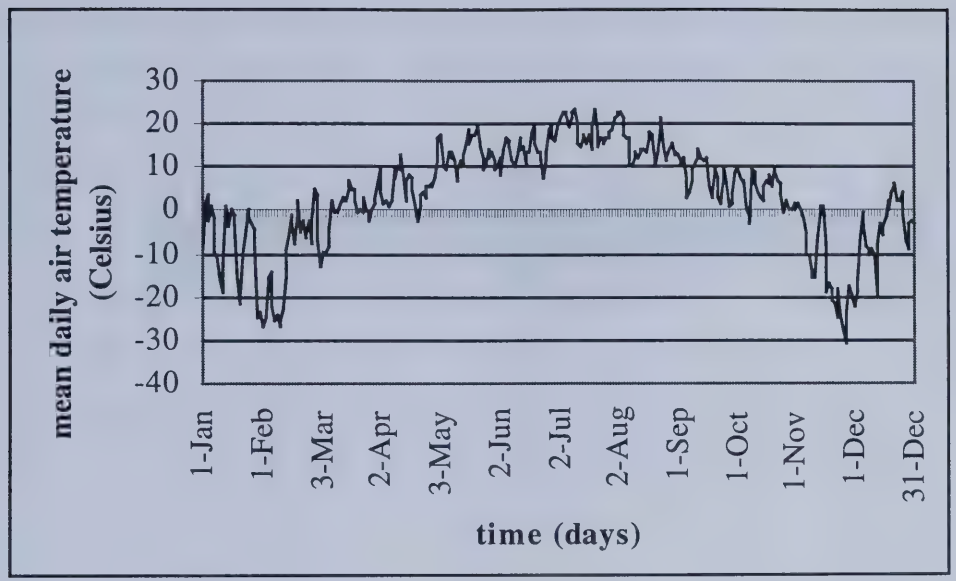


Figure 22. Average daily temperatures, 1985.

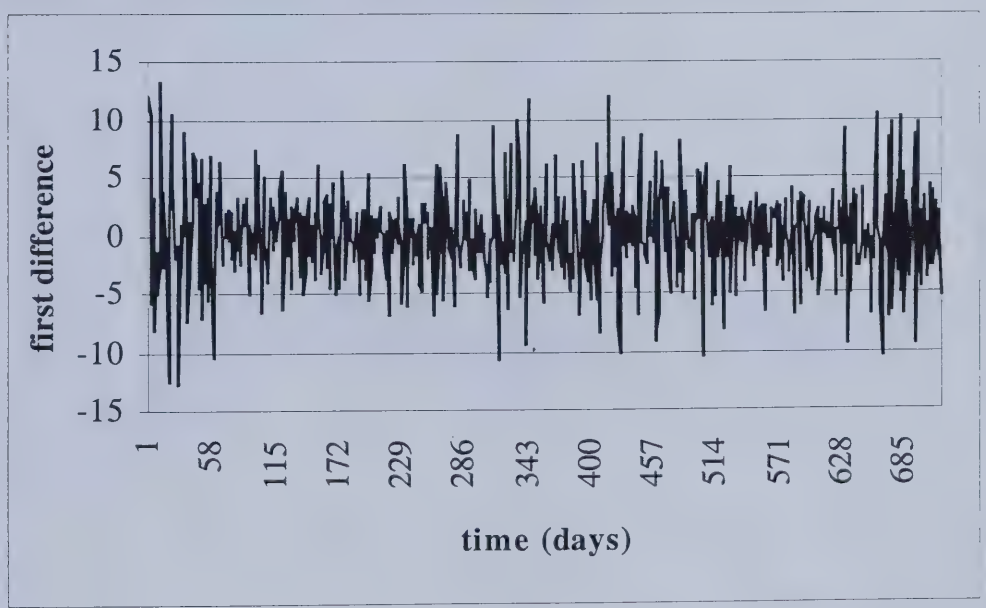


Figure 23. First difference of average air temperatures, 1985-1986.

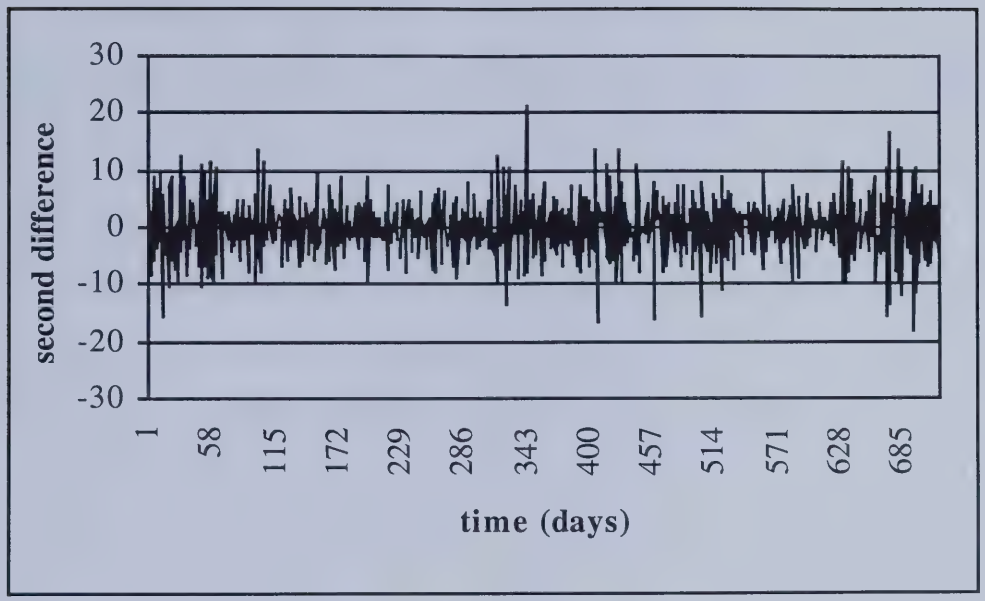


Figure 24. Second difference of average air temperatures, 1985-1986.

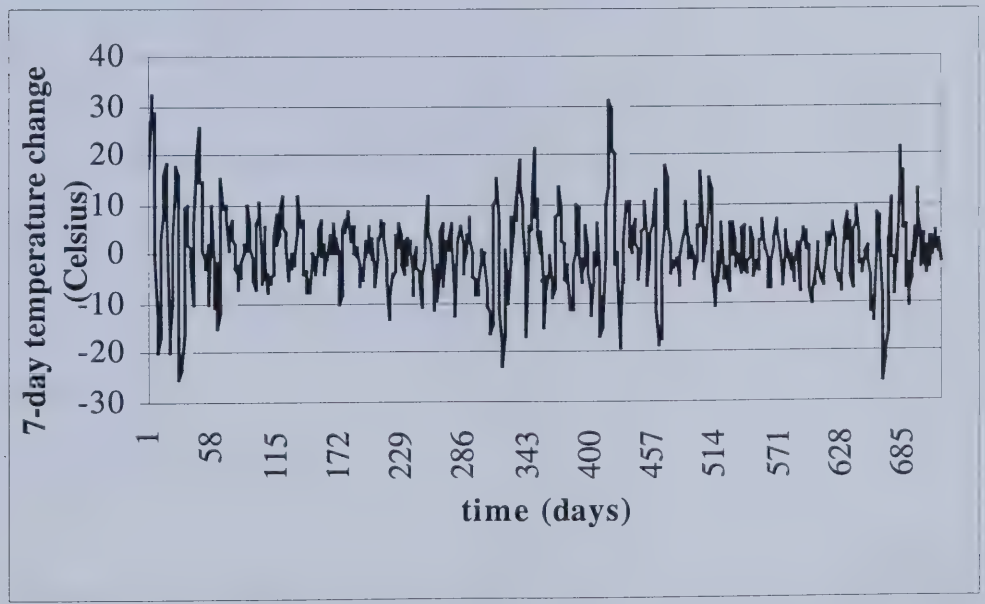


Figure 25. Seven-day air temperature difference, 1985.

Several alternatives were examined to determine which would best represent the different failure mechanisms acting on the pipelines. To imitate the major cold weather failure mechanisms (frost heave, soil-pipeline interaction, and water-ground temperature differential), multiple input parameters using air temperature were proposed. Due to the importance of cold weather, it was important that air temperature effects be accounted for in the model. As will be discussed in the model results and sensitivity analysis, the inclusion of multiple (modified) air temperature data as input parameters was justified in representing the different manners which air temperature affects pipe failure modes.

Modeling frost heave mechanics involved presenting the Artificial Neural Network model with input patterns designed to mimic frost heave rate. The resulting formula, presented in Equation 9 was used to calculate the frost heave rate characterization parameter. A daily time frame for measuring temperature changes was chosen to show how daily changes in air temperature translated to a frost penetration rate in the soil. This was calculated as follows:

$$\text{Frost Heave} = \text{Max}\left(t_{i+1} - \dot{t_i}\right) \rightarrow_{i=1}^{7} \qquad [9]$$

Where:

t = air temperature (°C)

Large consecutive negative differential temperature values indicated a faster rate of frost penetration. Fluctuating temperature differential values (alternating negative and positive values, or small negative values) indicated little or no change in the frost heave rate.

Large, positive consecutive values may indicate some warming trends, possibly having a ground thawing effect the number of consecutive periods is extended. This may result in differential ground thawing along the length of the pipe, resulting in more failures. A time lag was also inputted in order to facilitate memory by the ANN model for previously described extended periods.

Modeling of soil-pipeline interaction involved showing how sudden, significant changes in temperature, either during winter climate or transitions between warm and cold weather periods, resulted in pipe failures. To allow the ANN model to distinguish cold weather events and large temperature drops, two distinct measures were devised. The first parameter was a measure of the climate, warm or cold. The model was presented with a 7-day moving average of mean daily temperatures. The calculation for this was based on an arithmetic mean formula presented in Equation 10:

$$\text{Seven Day Average Temperature} = \sum_{i=1}^{n=7} \frac{t_i}{n} \qquad [10]$$

The second measure was a seven day change in temperature. This parameter was the maximum seven day temperature difference for the week previous to the data reference point, as illustrated in Equation 11.

$$\text{Maximum Seven Day Change} = \text{Max}\left(t_{i+6} - t_i\right) \quad i = 1 \rightarrow 7 \qquad [11]$$

The purpose of this parameter was to indicate a contraction of the pipe, causing an increase in tensile stress along the length of the pipe. As described in the soil-pipeline interaction section (as a failure mechanism), this type of action may result in pipe failures.

Internal Pipe Water Temperature

The possible sources of pipe water temperature were investigated. The initial source investigated were the Rosslyn reservoirs' water temperatures, as this was the nearest water reservoir to the Calder area, and therefore was the most reflective of the water temperature within the water pipes. This data was analyzed for its reliability. Graphical analysis, as illustrated in Figure 26, revealed inconsistencies in trends, which would present "noise" for the ANN model, thus reducing its accuracy.

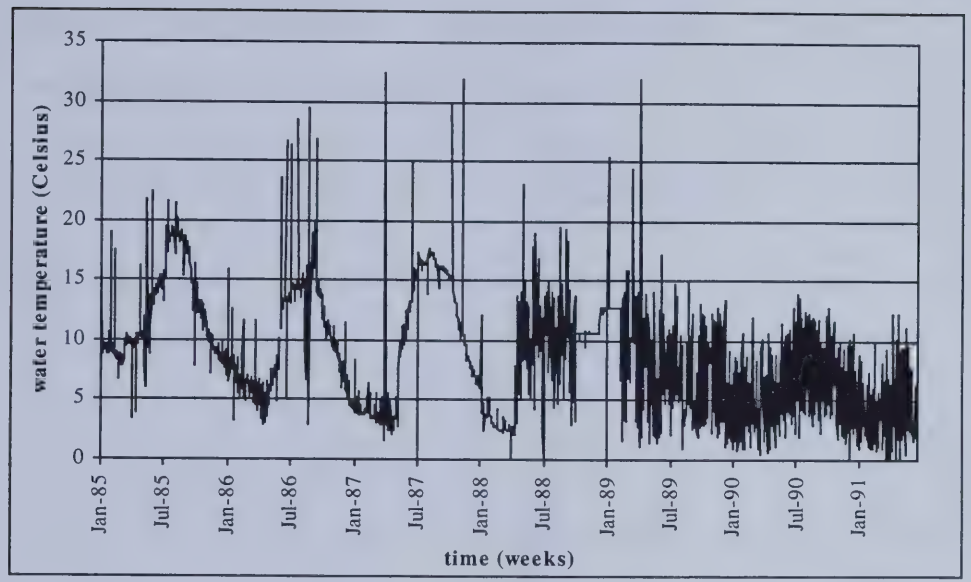


Figure 26. Graphical analysis of Rosslyn reservoir average water temperature, 1985-1991.

Referring to Figure 26, the data appears to have a well-defined trend from 1985 until 1988 after which, the trend changed, appearing more erratic. This was discussed with Mr. Ken Richardson, supervisor of Water Transmission at Aqualta, and it was decided that this data was not reliable. The reason was that the reservoir sensors did not detect water temperature during station shut downs (lasting for weeks), but instead measured the room temperature. Thus the sensors did not measure water temperature unless the water was running. As a result of these inaccuracies, a surrogate measure using the water treatment plant intake temperatures was investigated. It was concluded that the water treatment plant water temperatures from which the reservoir water originated could be used as surrogate data. It was felt that this data was reliable since detention time in the

reservoir was not significantly long enough that the water treatment plant's temperatures would not be would representative of reservoir water temperature. The seven-day average water intake temperature is shown in Figure 27.

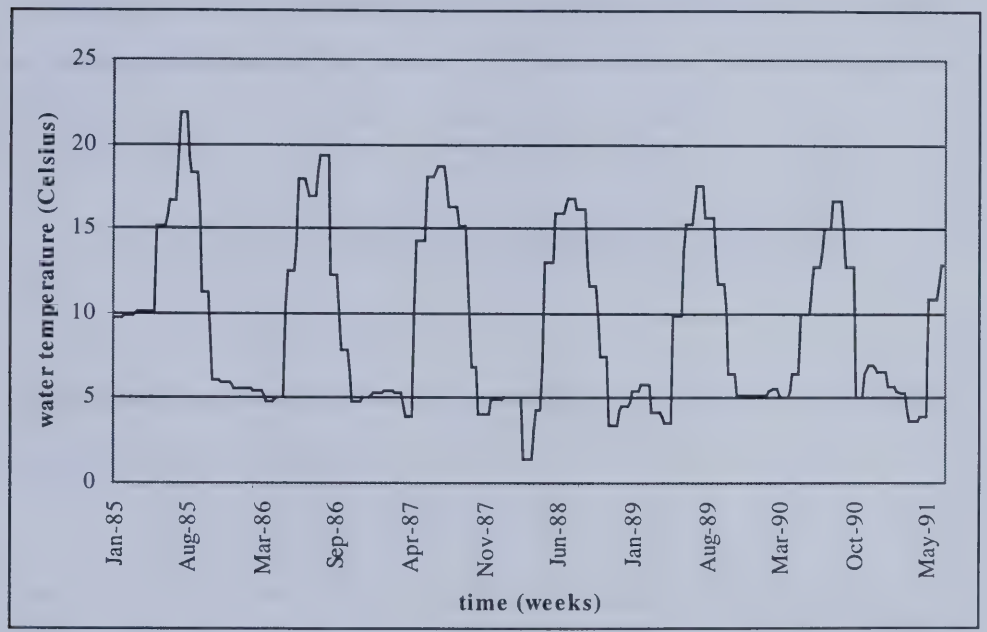


Figure 27. Seven day average water temperature (Rossdale Water Treatment Plant) 1985-1991.

Because the thermal conductivities of soil and water are different, meteorological fluctuations result in different rates of transition for water and ground temperatures. Drops in ambient air temperature following periods of relatively constant temperature affect the thermal exchange rates from air to water and to soil in differing degrees. This results in a substantial temperature difference between the ground surrounding the pipe, and the water within the pipe. This difference creates a temperature gradient across the thickness of the pipe wall, resulting in differential strains and longitudinal stresses. This hoop stress condition increases the likelihood of longitudinal failures (Habibian, 1994).

Thus internal pipe water temperature, coupled with an indicator factor of ground temperature would help physically represent the occurrence of hoop stress conditions.

In order to model pipe wall temperature differential stress, an expression for magnitude and temperature differential was required. This was calculated by the following formula (Equation 12):

$$\text{Seven Day Average (Air - Water) Temperature Differential} = \left(\sum_{i=1}^{n=7} \frac{t_i}{n}\right) - \left(\sum_{i=1}^{n=7} \frac{w_i}{n}\right) \qquad [12]$$

This input parameter would properly demonstrate instances where hoop stresses were probable to occur.

It was thought that expressing a pipe wall temperature gradient exclusively might be incomplete because it did not reflect other stress conditions occurring concurrently. The presence of contributing stress conditions may be necessary for this type of failure. Because of this possibility, a magnitude indicator in the form of the 7-day average water temperature was added. It calculated in the same manner as the 7-day average air temperature. This expressed the overall temperature conditions. High pipe wall temperature differential events occurring at sub-zero temperatures would be much more significant than those occurrences at warmer temperatures, speculatively due to decreased pipe ductility with decreased temperature.

Operating Pressures

The main concern with operating pressure as a parameter is the effect of unexpected events such as pump shutdowns and other types of pressure-related problems. Water hammer effects resulting from pump shut downs and accidental valve closures can result in blowout failures, especially on pipe walls which have been weakened by corrosion effects or previously stressed pipe. Failures may also occur due to operating pressure, which are not within the design operating range for the pipe.

Investigations into the availability of operating pressure data showed that only average daily pressures were recorded. For this reason, it was not reasonable for predicting failures occurring from water hammer effects, since these are instantaneous events, and seldom are accurate records kept. However, the benefit of using this parameter was not completely discounted. The average seven day operating pressure was included as a potential input parameter to determine if long-term operating pressures have an effect on the failures of the mains.

Pressure information was gathered from reservoir data logs. The water pressure input parameter was simplified by assuming that the pressure from the reservoir servicing the subdivision would be an adequate indicator of the area's overall average pressure. Further manipulation of data would have required specific knowledge of the elevations of various points within the service area, and calculating a hydraulic grade line. This manipulation was thought to be unnecessary since the reservoir area was static, and

therefore the same reference point was always used. In addition, due to the generality of the modeling (probability of pipe failures is only being predicted for the Calder subdivision as a whole unit, not for individual pipe breaks), this level of information was considered unnecessary.

Rain Precipitation

Rain precipitation is an important parameter when considered in relation to soil type. The Calder area is composed entirely of a silty clay loam. Clayey soils will be affected to a larger extent than sandy soils when there are periods of low precipitation followed by periods of high precipitation. The inclusion of this parameter may be especially useful for prediction of failures during the warmer seasonal months, as opposed to the cold-weather input parameters. This was a distinct possibility after graphically analyzing the pipe breaks, since not all of the water main failures occurred during the colder months, shown in Figure 28.

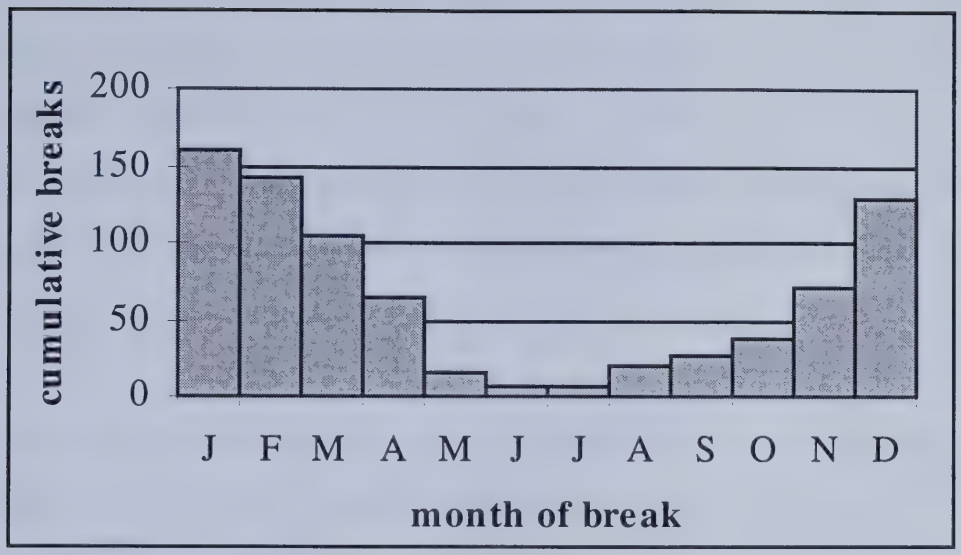


Figure 28. Total monthly breaks, 1985-1991.

Several time periods were correlated in initial models, to determine which time period would best model the physical process. Results favored the inclusion of a 180-day moving average. This data was obtained in conjunction with the air temperature data.

Cluster Indices

Historical studies of water main failures in other cities show that pipe failures will often occur in clusters, as demonstrated by studies conducted in Winnipeg, Manitoba [(Goulter and Kazemi, 1988) and (Goulter and Kazemi, 1989)].

This phenomenon may be caused by a disturbance of soil surrounding a pipe break, due to the pipe failure (i.e. water seepage causing instability of the soil), or even the repair of

the pipe failure (excavation and replacement of soil causing disturbances to soil strengths). Regardless, the studies demonstrated that this type of phenomenon was significant. Due to the relative proximity of the failures (both in time and space), they might be mistaken for a multiple break event, when in actuality there would be multiple causes (i.e. the primary failure cause, and then the resulting additional failures from water seepage or other causes).

Due to the large area selected for modeling, it was impossible to input a spatial cluster input parameter that would correspond to the initial break. This was due to spatial variability of pipe breaks over the entire area, and in this case the model output was a summed output. Therefore the model was not designed to determine the exact location of a pipe break, and a spatial cluster parameter was not a truly indicative input parameter. For example, trying to spatially correlate two separate pipe failures, several kilometres apart, (under the assumption that one failure is linked to the other) would be incorrect, causing noise and possibly decreasing model accuracy.

Although a spatial cluster parameter was not thought to be useful, a temporal index parameter was created using a one-week lag of breaks within the study area. In this manner, it was thought that pipe breaks occurring in the time period immediately preceding the current period would be able to adequately represent this phenomenon. Therefore, the previous week's number of breaks was inputted to show temporal correlation. It was decided that due to the time frame selected (one week intervals),

periods longer than two weeks (the present week interval and the week before) would not be physically meaningful for temporal correlation.

Pipe Integrity

The pipe integrity parameter was seen as an important indicator of the overall structural integrity of the water mains within the subdivision. This measure was calculated by summing the total number of pipe breaks that occurred in a one year span previous to the input data point in time. It was thought that this input parameter would indirectly infer whether these pipes were affected by corrosion and would give an approximate measure of the potential degree of this corrosion. This was important since soil parameters indicative of corrosion were not available in sufficient detail or quantity to be useful for model development. The number of breaks of the same pipe diameter within the previous year was accepted as a correlated indicator of pipe integrity, as greater pipe integrity would be indicated by fewer breaks within the previous year. Conversely, more breaks would indicate a more weakened overall state of pipes within the study area, and thus higher pipe failure frequency.

This parameter was a moving total value, and each input pattern included only the total number of breaks within the previous year. Analysis of the trends showed that the total number of breaks within the previous year fluctuated (shown in Figure 29), which seemed peculiar. It would be expected that the total number of breaks within the previous year would continually increase, to reflect the continually worsening degree of

corrosion. However, several factors may explain why this was not seen. Variant weather patterns, i.e., severe or mild winters may effect yearly break rates. Effective mitigation techniques, e.g., cathodic protection, extensive pipe repairs, and pipe replacements may be another factor for variation. Also, random and/or exceptional events, those severe enough to cause additional failures, may be indicative of an uncharacteristic year. Therefore, historical information is intrinsically embedded in this input parameter.

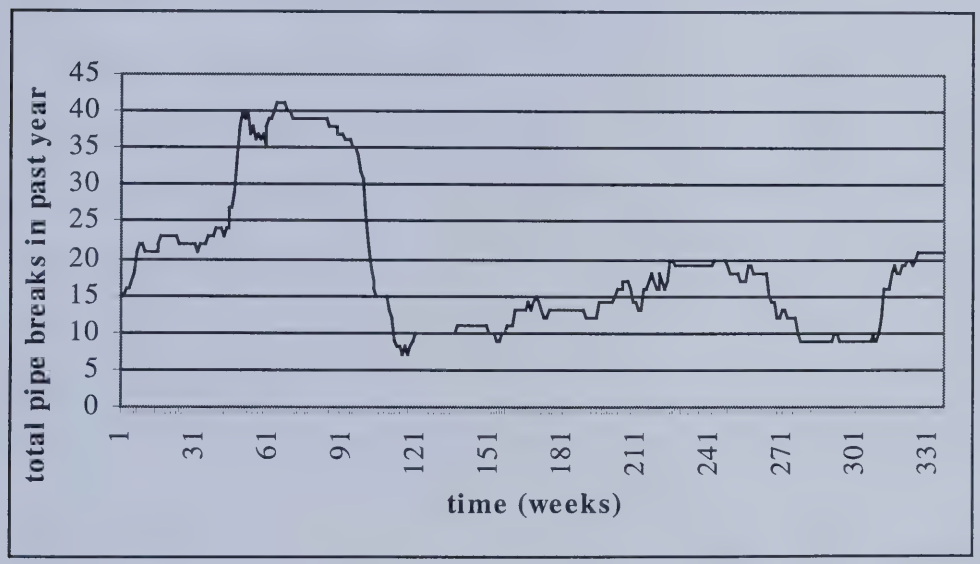


Figure 29. Previous year total pipe breaks (moving total), 1985-1991.

Omission of Potential Input Parameters

From the above description, it becomes obvious that a number of causal or influencing factors (from literature, summarized in Table 2) have been transformed and/or replaced with surrogate input parameters. Others are omitted altogether. Those parameters selected as potential model input parameters are summarized in Table 4. Lagged values

for each component (i.e., the parameter's value of the week directly previous) are indicated if they were considered potential input parameters.

Table 4. Summary of Potential Input Parameters.

Input Parameter	Lag Component
Air temperature: maximum daily difference	Yes
Air temperature: max absolute daily difference	Yes
Air temperature: maximum 7-day change	Yes
Air temperature: maximum absolute 7-day change	Yes
Air temperature: 7-day average	Yes
7-day average water temperature	Yes
7-day average air-water differential	Yes
7-day average pressure	Yes
180-day total precipitation	No
Previous week lag	N/A
Previous year lag	N/A

Factors identified in literature relating to pipe failure mechanisms or indicative of the durability of the individual pipes were not included. This was because of the manner in which the study area was modeled. The individual pipe's overall physical characteristics, such as pipe diameter and type of pipe joint connections, were not included because of the difficulty in transforming these individual pipe characteristics into an overall area characteristic (certain areas were installed years later, others had been replaced). This

was the primary reason for the scope of the study to be limited to one size of pipe. It was also seen as advantageous, as it is easier, and also more accurate to model a single output.

Environmental parameters affecting the pipe, such as overburden pressure, soil type and properties, soil pH, and soil water content were omitted due to lack of monitoring of the data. Review of literature showed these parameters to be appropriate measures of corrosion and soil-pipeline interaction strengths. However, because the purpose of the model was to predict the likelihood of failures for an overall study area, and for prediction of cast iron water mains only, the above parameters were presumed adequate. Models were developed utilizing the above input parameters.

3.1.2 Open-Domain Problems

Modeling of the Calder area pipe failures had difficulties because it is an open-domain system. An open-domain is a system that is not controlled, and in which interferences such as alternate causes, can neither be identified nor quantified for their overall effect within the system.

As a result of working with an open-domain system, model development using Artificial Neural Networks was the best methodology. Nonetheless, within an ANN-type model, modeling difficulties were anticipated. Exposure to a full range of data patterns in time series was required for the ANN model to make appropriate input-to-output parameter associations and, therefore, accurate models. Conversely, a lack of appropriate data,

including boundary conditions, is extremely unfavorable for developing accurate deterministic-type models. A deterministic model typically requires a controlled, closed-domain system. This type of model will derive values based on magnitudes of measures, and often not in time series. Therefore, ANN methodology has an advantage in open-domain studies since it does not necessarily require boundary conditions.

3.1.2.1 Time-Frame Selection

To contend with the open domain nature of pipe breaks, a suitable time frame was investigated. To reasonably model the pipe break process, it was essential to ascertain this time frame for which the input parameters could correlate to the output parameter (pipe breaks). Initial trials with daily data were attempted, but it was found that some input parameters, primarily those temperature-related, could not adequately form an association with pipe breaks. This could be attributed to the time required for cold air temperature effects to propagate into the ground to pipe depth or water. A daily interval also did not allow the ANN model to correlate an air temperature drop with a pipe break. Model results using an arbitrarily set seven-day interval were reasonable, and greater accuracy in prediction was immediately encountered. Due to the association with the seven-day interval (equating to a weekly interval, as opposed to using a four-, five- or six-day interval) interpretation of the data in a physical time sense was also more easily comprehended.

Weekly time intervals were implemented in the model by taking seven-day averages, seven-day totals, or maximum daily values within the one-week interval, depending on

each parameter's function in the model. Where average values were required, daily input data points were calculated using the reference date, and the six days previous to this date for averaging purposes. For instance, averaged values from January 1st to January 7th would be included in the January 7th reference data point. Where seven-day totals were required, the same principal was used. For the same example, a total of the daily values from January 1st to January 7th (inclusive) would be included in the January 7th data reference point. For maximum values, the maximum daily value within the seven days would be selected for modeling.

3.1.3 Limitations

As discussed previously, the plausibility of developing a successful model hinges on the quality of the data available. It was demonstrated in the literature review and source data analysis that the largest limiting factor for modeling ease and accuracy is the presence of appropriate and comprehensive data.

There was a discrepancy between ideal input parameters and the actual data available. This discrepancy serves to emphasize the point that in order to facilitate ease and accuracy in development of an Artificial Neural Network model, limitations to output and input parameters must be overcome.

3.1.3.1 Limitations of Output Parameter

A proposed model output parameter employing an area break density output (number of pipe breaks/km of pipe/km^2 of ground area) was investigated instead of a linear break density. This type of output was thought to be more valuable than a linear break density since it would have the ability to define and pinpoint a smaller area for use in Edmonton's water main replacement program. However, information of this type was not readily available, and obtaining the parameter in this form required meticulous manipulation of spreadsheet data and city maps. Much of the historical data could not be found in spreadsheet form, although there is current work ongoing attempting to update city databases. Therefore, the linear break density was adopted.

3.1.3.2 Limitations of Input Parameters

The availability and reliability of the raw source data created the need for extensive analysis and manipulation of the raw data in order to provide input data patterns representative of the failure mechanisms. The amount of analysis (and re-analysis required) hindered the timely development of the models.

Because of the open-domain nature of the system, and its effect on the correlation between input and output patterns, unpredictable results could be expected. Dealing with these problems involved making assumptions. Fair assumptions could be made, supported by theory (from literature) or by circumstances (small areas may be assumed to

have similar characteristics). However, this will tend to limit the range of application unless the availability and reliability problems are mitigated.

Soil type, an important factor for several failure mechanisms, was not used in the model because available soil maps were too general, and indicated the Calder area was composed entirely of a silty clay loam, as was a great majority of the Edmonton area. Finally, due to a lack of monitoring of most of the soil parameters mentioned, it is impossible to include any such information without taking samples from individual areas. This process would be time consuming and not economically feasible. For this reason, the developed model is limited to areas where uniform soil conditions can be assumed.

3.2 Modeling Methodology

The four-step model development methodology for Artificial Neural Network was generally followed. As mentioned earlier these four steps are: Source Data Analysis; System Priming; System Fine-Tuning, and; Model Evaluation. As will be demonstrated, the Source Data Analysis stage has already been defined and performed in the above section. System priming has already been partially completed, and was done concurrently with system fine-tuning. These middle stages involved employing systematic manipulations of data and ANN model structures to arrive at the best models. Model Evaluation is the stage where performance criteria are defined and best models chosen.

In source data analysis, the first objective of this exercise was to understand the problem being modeled and to establish the cause-effect and influencing factors as they pertained to the output. Having done this through careful research and review of available literature, it would be possible to obtain the necessary data. The second purpose of the source data analysis stage was to determine the reliability of the data, and to prepare the data for input into the ANN model. Given that the study domain modeled was an open system, uncertainty of the inputs was unavoidable. Inaccuracies in measurement of input parameters (instrumentation accuracy and tolerances may not always be good; many parameters are not measured with satisfactory frequency) resulted in "noise" during the training of the model, possibly affecting the model's forecasting capabilities. For this reason, it was extremely important to have input parameters that would accurately reflect the physical pipe failure mechanisms.

As illustrated in the above section, there were a number of instances where the availability or reliability of the raw input data collected imposed limitations. Therefore inputting the most representative patterns and use of suitable ANN model structures would be paramount for accurate model development. This was achieved in the system priming and system fine-tuning steps.

In the System Priming stage, the overall objective was to determine which of the potential model input parameters would produce the best predictive model. This involved inputting patterns which best described the cause-effect and influencing factors. Having done this, the next stage involved methodically determining which of these

parameters should be included in the model. Again, for open-domain system studies, this process becomes more difficult because of the need for confidence in the relevance and reliability of the data.

Due to the use of a number of transformed and surrogate input parameters, it was decided that the system priming and system fine-tuning steps would be developed interchangeably with one another. This reasoning became evident during the model training stage. Early results did not indicate satisfactory prediction capabilities, despite use of sound methodologies. Therefore an iterative process was employed between the system priming and system fine-tuning stages. The methodologies applied will be further described in the following section. Further model refinement and input parameter inclusions or exclusions would be decided upon depending on the iterative modeling results.

The System Fine-Tuning stage consisted of meeting three goals. The first, was to determine the most appropriate way to present the data sets (training, test, and production sets) to the ANN model in order to allow the model to "learn" the data by appropriate associations, without having it "memorize" the data. The second goal was to distribute events within the data sets such that the full range of occurrences would be captured by the model. These first two goals were extremely important to maintain the forecasting power of the model. The third goal of system fine-tuning was to develop the model type and model structures which would most easily and efficiently model the domain, while including features typical for the process.

As previously mentioned, the model structure, the distribution of data sets and data points, and model input parameters were determined in conjunction with one another. Determining the best combination of these elements involved iterative trials, however use of analytical modeling methodologies allowed for the most efficient convergence. Details of the different methodologies employed are to be discussed in this section.

3.2.1 Model Progression

For preliminary modeling, a wide range of model architectures were examined (Standard connections; Jump connections; Recurrent network and Ward networks). This was performed to determine if any particular network was more advantageous for the inputs chosen. The most important potential input parameters were inputted into each model type. The models were then evaluated by the R^2 statistic (a measure of the model's predictive error when compared to actual output data). This measure was applied to the cumulative ANN model predictions. This method allowed for a fair, overall comparison of the prediction errors. A secondary evaluation was by visual inspection of actual versus predicted results, to surmise whether pipe break trends were being matched. Based on these results, it was evident that the standard connection backpropagation networks were the most suitable network architectures. The standard networks, using 1 or 2 hidden layers, provided the greatest potential for model development. The standard connection network was also desirable for preventing complication of the model

development process and to allow for further application of the model, since this model architecture type is quite simple in structure.

Once the general model type was determined, the next task was to determine which model input parameters would be most relevant for model accuracy, and in what proportions these data sets should be presented to the model. A portion of the data was selected for the training and testing sets, and the remainder was used in the production set. The predictive properties of the model were then tested on the entire data set (the pattern file, consisting of the training, test and production sets). This pattern file was arranged in chronological order to maintain the time-series predictive properties of the model.

To determine each input parameter's significance to the models, a number of methods were attempted, using the R^2 statistics of the production and pattern files. An addition-type method was used. This method involved gradually adding more input parameters, starting with five input parameters. A factorial design procedure including all parameters was also utilized in order to determine the appropriate input parameters. However, the significance of individual parameters was not readily evident through this type of analysis since parameter interactions were significant. A substitution-type procedure was then employed. This method involved removing a single parameter to determine the parameter's significance. This parameter was then replaced, and a different parameter excluded. Results from each method were best using addition- and substitution-type models.

Different configurations of the standard backpropagation networks were tried concurrently with the determination of the important input parameters. Due to the interdependence of the input parameters and the type of network model used, a large number of trials were attempted. In each case, the model structure was kept constant while the determination of the important input parameters proceeded. Once progress was made in determining relative importance of an input parameter (depending on the method of input parameter determination being performed) a change in the network configuration (i.e., number of layers, number of neurons, ratio of neurons in adjacent layers, etc.) was performed. This process was iterated, until a large number of models with calculated cumulative R^2 statistics (on the models' production sets and pattern files) were accumulated. From the above, ten overall potential models were selected.

From the selected potential models, minor manipulation of the data set proportions and data points were performed. The purpose for doing this was to further optimize the potential models, and ensure the selected models were exposed to the full range of data patterns. These model variations were compared using the R^2 statistic applied to the models' pattern files and the two best models chosen for further manipulation. As a final step, the activation functions were also varied. Lastly, final model evaluations were made.

Model evaluation is the selection of the developed model that demonstrates the overall minimal error in prediction, while correctly predicting the output trends based on the

given information. A criterion for this selection was the best combination of R^2 value and visual inspection (to match trends) indicated by slope comparison of actual and predicted break values. Consideration of the models' predictive robustness (sensitivity to changes in parameters—evaluation of model logic, and if the model appropriately accommodates changes)was also factored. The best models are described in Section 4.0.

3.2.2 Data Set Manipulation

When training the Artificial Neural Network model, it must be realized that accurate models will only be achieved if the manner in which the data is presented is appropriate. Therefore, it is important that a representative amount of the potential data is exposed to the model for training and testing of the model's progress. Without having this representative range of historical events to detect and verify, the model will not be able to predict events too far outside of the learned cause-effect logic.

Determination of the optimal amount of historical data required for models' training, testing and production sets was based on a trial-and-error method. For a given model (one with established predictive ability), the fraction of data to be inputted into each set was varied. Results from the R^2 and trend predicting ability were analyzed to determine the effect of varying the data set proportions. Too little information in the training and testing sets indicated poor results in the production set. This may be likened to poor learning of the model. Too much information in the training and testing resulted in a comparable decrease in trend prediction and R^2 value when applied to the production set.

This may be likened to the model memorizing only the information presented to the model, and then performing poorly when attempting to apply itself to new data patterns with slight differences.

Division of input data into training, testing and production sets was performed on a number of different selection criteria in order to determine if any of these methods was more time efficient in terms of training time. These included:

1. Random selection;
2. Frequent Interval selection;
3. Yearly interval selection;
4. Specific data point selection (based on sorted output results);
5. Specific data point selection (based on sorted output and grouped input division); and
6. Event selection/elimination (based on model results).

Random selection refers to a semi-random division performed by the *Neuroshell 2* program. This is considered semi-random since the program requires the input of a "seed" value, in order to divide the pattern file into the three data sets (test, training and production). Input of an identical seed value results in identical division of the same file. *Frequent interval* selection acts very similarly, except the modeler specifies the frequency of selection of a pattern into the three data sets. These selection criteria are the most automatic with little adjustment or judgement from the modeler.

The remainder of the selection criteria required use of judgement by the researcher. The yearly interval selection criteria was based on the idea that the data must be presented to the model in time series, since the output was determined in time series. Specific data point selection, where individual data patterns, or small groups of data patterns (up to five patterns) were selected for inclusion into one of the three data sets. Selection of data patterns involved maximizing the exposure of the output and inputs in all data ranges, based on mechanical sorting of the values of the output parameters alone, or mechanical sorting of the data patterns in the input and output parameters at once. Event selection/elimination employed the GRNN models to determine how an individual data pattern effects the prediction accuracy. This is described below.

3.2.3 Data Point Manipulation: Selection/Elimination Protocol

By training the models using this method of data set division, the methodology involved evaluating individual data patterns, or a small set of data patterns (five or less) in the pattern file. A decision was then made whether the individual data pattern should be included in the training, testing or production set. This decision-making process employed the following protocol:

1. R^2 model improvement;
2. False prediction of non-break events (phantom peaks) and;
3. Prediction of break events

The GRNN model was used because of the learning architecture's ability to memorize input patterns when presented with sparse data sets (Ward Systems Group, 1993).

The reasoning for this decision-making protocol was to employ a supplementary teaching method to improve the models' exposure to specific events. It allowed the ANN model to be exposed to an event and acquire the appropriate correlation to the output based on recognition of the specific input pattern. In essence, the model had to be tutored, as only part of the logic of the problem was being originally learned. Therefore, the above learning method allowed the model to become accustomed to learning particular data patterns.

The following two figures demonstrate the effect of moving five data patterns from the pattern file into the training file. In this example, the model was initially trained using only break events through a GRNN architecture (see results in Figure 30).

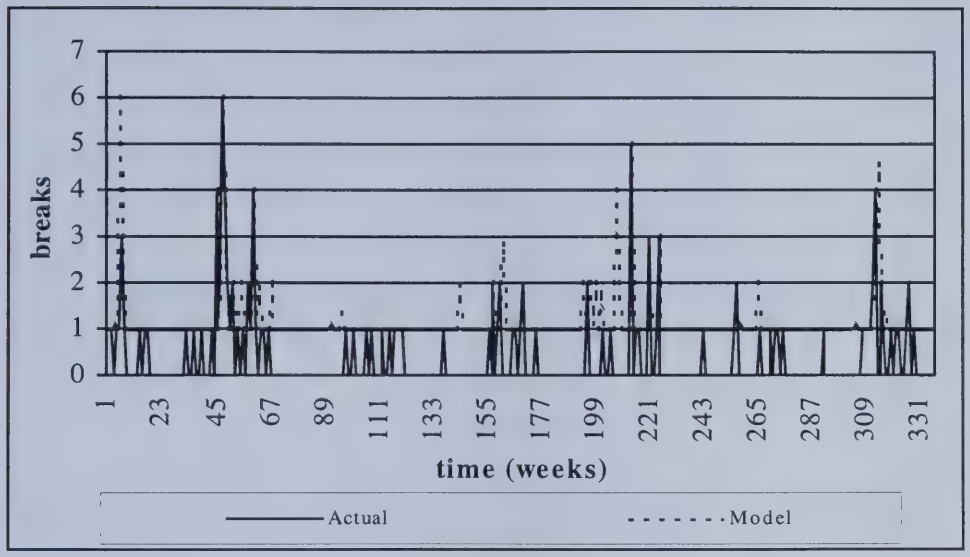


Figure 30. Event prediction before training with non-break data.

Subsequently, five data patterns were selected from the pattern file, and distributed to the training and testing sets. The result of this manipulation on model results is illustrated in Figure 31.

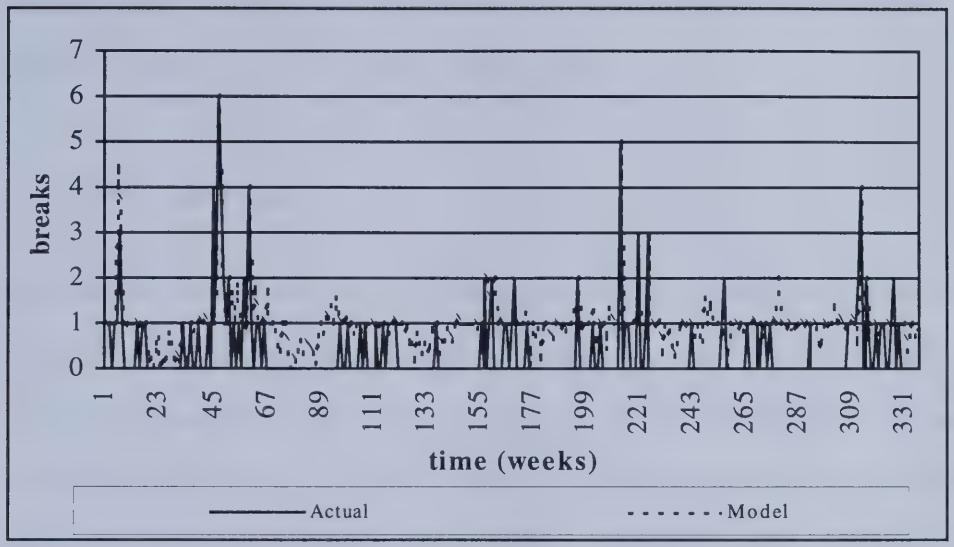


Figure 31. Event prediction after "point A" data point addition

From these results, it is demonstrated that there is a potential use of this data selection technique. While this technique requires judgement, and time to evaluate the results, it has useful applications for studies such as this, when studies are of open-domain systems and data is limited.

3.2.4 Input Parameter Selection: Methods Tried, Results

Three distinct methodologies were employed to determine the appropriateness of the potential input parameters for the pipe break study. Initially, a methodology was applied which involved gradually increasing the number of potential input parameters, and then determining the effect on the ANN model. A factorial design procedure was also attempted, to take into consideration the significance of the interaction between the input parameters. Thirdly, substitution of individual input parameters was attempted, to

determine the effect of the input's absence on the model. Evaluation of the different parameters was based on the R^2 statistic of the model.

3.2.4.1 Inputs Addition

Initial neural network models were configured by varying a small number of parameters that were thought to be the more important parameters. The input parameters which were initially investigated were: 7-day average temperature; 7-day temperature change; 7-day average water temperature; (air-water) temperature differential; and total pipe breaks for one year previous. It was anticipated that significant changes in modeling accuracy could be observed in order to predict major trends. A second goal was to determine which form of the input parameter (specifically air temperature) was most appropriate for modeling accuracy.

This method involved inputting initial models with only the parameters that were thought to be most important, as per cited literature. By beginning with only the bare minimum number of input parameters, and then gradually adding more potentially input parameters, the models being developed would gradually increase in R^2 value. The addition of those input parameters that provided marginal or no improvement were not significant to modeling the study and, therefore, were excluded

3.2.4.2 Factorial Design

Factorial design analysis of the potential input parameters involved selecting a fixed network followed by varying the parameters that were likely to be redundant or having significant interaction effects. Evaluation of this method involved factorial analysis of the production and pattern file R^2 values. The R^2 of the production file was used to determine the model's learning ability of the patterns, while the R^2 for the pattern file was used in order to maintain time-series prediction.

Care was taken so that a small enough range for each input parameter was chosen. Because of the inherent non-linearity of the network configuration, the effects of varying a single factor was also non-linear. Therefore, if due care was not taken, it was possible that a change in the effect would not be noticed. Also, because of the non-linear behavior, it was necessary that either the number of input parameters or type of input parameter could vary depending on the type of model chosen (e.g. Recurrent Networks compared to Standard Networks). Due to this non-linear variability, an iteration process (checking fit of various model types against varying input parameters) was required, creating the need to perform several hundred model runs.

3.2.4.3 Substitution-Elimination

This process involved removing and replacing a single potential input parameter, and then evaluating it's effects on the R^2 statistical value. A lower value of R^2 indicated

significance of the parameter. A higher R^2 statistic indicated no significance or noise in the parameter. This process may be considered a trial-and-error type experimentation.

4.0 RESULTS

4.1 Evaluation Criteria

Evaluation of the potential models involved a combination of three factors: R^2 model fit statistics, trend prediction ability (slope matching) and simplicity (model architecture complexity, measured by the number of hidden layers, and total number of neurons). Reasons for this multiple criteria were that the best model must have the ability to:

1. Predict the events with accuracy (R^2 value);
2. To have a strong trend prediction (the rate at which breaks would occur, at any particular time frame) and;
3. To be simple enough to be used for further modeling purposes.

Based on these criteria, the best models were chosen. Evaluation was based upon a combination of quantitative measures and good judgement, since only the first criterion is completely quantitative. Trend prediction is subjective, since slope changes are so rapid and frequent, and visual slope matching (of actual versus predicted trends) for particular time periods is most important. Simplicity of the model is also subjective, and depends on the expertise of the modeler.

4.1.1 R^2 Statistic

Calculation of the R^2 value for the cumulative model was performed using the following formula (Equation 13):

$$R^2 = 1 - \frac{SSE}{SS_{yy}} \qquad [13]$$

Where:

$$SSE = \sum (y - \hat{y})^2$$

$$SS_{yy} = \sum \left(y - \bar{y} \right)^2$$

Where:

y actual value

$\hat{y}$ predicted value

$\bar{y}$ mean value of y

This was calculated using the applied pattern file since cumulative breaks is not actually calculated in the actual ANN model. Instead, EXCEL was used to sum up breaks in the time-series, and then used (according to the above equation) to calculate the R^2 value. The R^2 statistic depicted in model results is applied to the entire data set (6 ½ year of data). This statistic indicates an overall accuracy of the model, such that the probability of events occurring is predicted well on a regular basis.

4.1.2 Trends Prediction

Trends prediction ability was based on matching the slope of the actual results and model prediction results. Overlapping or parallel lines indicated good forecasting ability since break rates were matched. Gaps between the two lines are not indicative of trend predictive ability since any errors in prediction are accumulated throughout the length of time the model is being evaluated. These gaps are indicated in the R^2 model fit statistic.

4.1.3 Model Simplicity

The importance of simplicity of the model cannot be underestimated. Maintenance of a simple model is paramount for reproducibility of results as well as implementation of the model. Without this simplicity, future models may become convoluted with unnecessary parameters and the importance of these parameters will not be understood.

4.2 Best Models (Cumulative Results)

Determination of the best model was based primarily on the models' cumulative R^2 values. This measure gives a good overall indication of the models' predictive capabilities by showing overall errors instead of focusing on single-event accuracy. Ten prospective models were selected based on the R^2 criteria. These models were further

evaluated on their trend-predictive ability, with consideration given to the simplicity of the models for future application.

The two best prospective models are shown below. Model A demonstrated excellent predictive ability employing only 5 parameters. Graphical results are illustrated in Figure 32, with model properties summarized in Table 5. Model B demonstrated even better trend-predictive accuracy with 13 parameters. This model's results are shown in Figure 33 and summarized in Table 6.

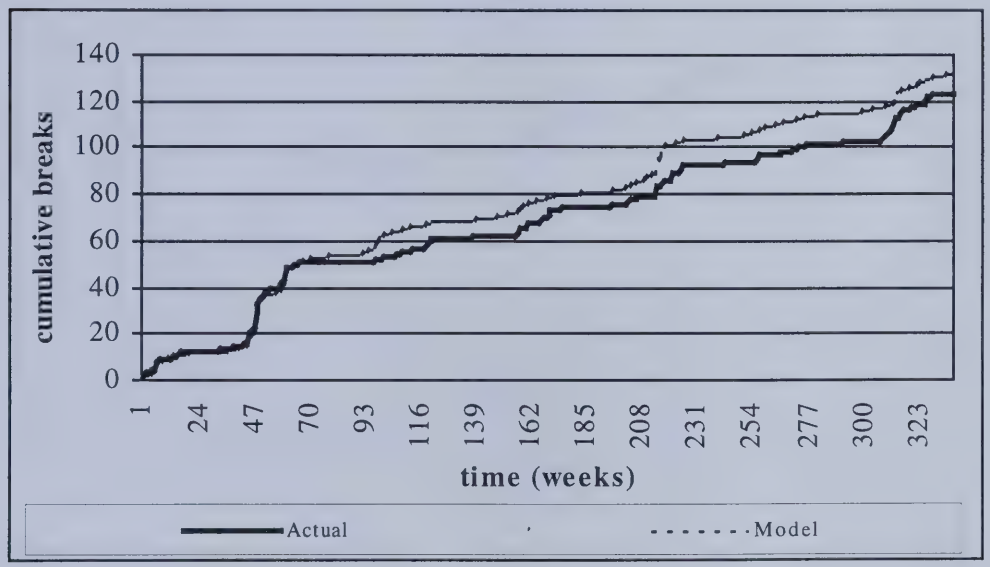


Figure 32. Prospective model A.

Table 5. Prospective model A: model specifics.

R^2 value	Network Architecture	Configuration	Activation Function
0.920	Standard Backpropagation	5-25-6-1	Linear [-1,1] (scaling) Logistic (all other)

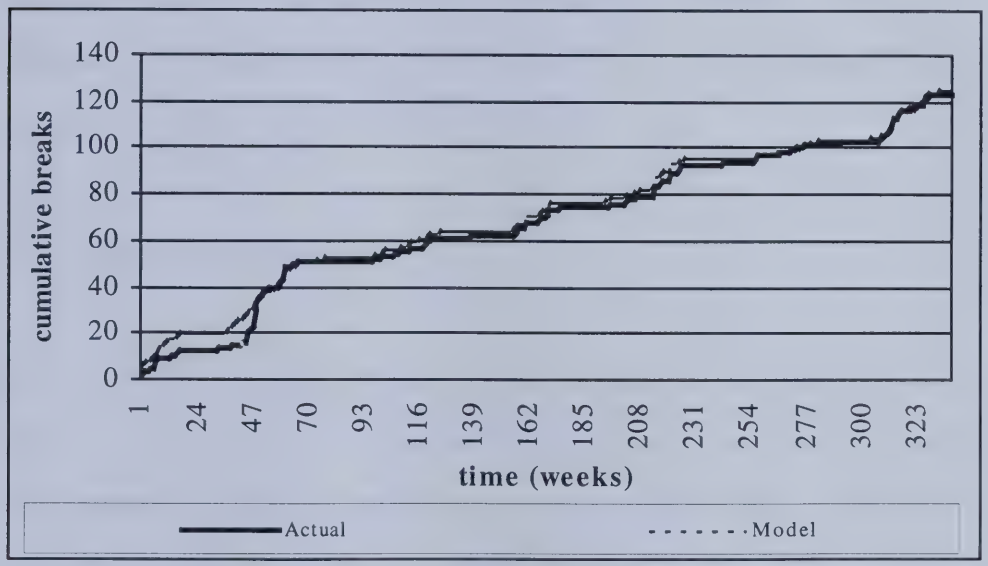


Figure 33. Prospective model B.

Table 6. Prospective Model B: model specifics.

R^2 value	Network Architecture	Configuration	Activation Function
0.986	Standard Backpropagation	13-39-1	Linear [-1,1] (scaling) Logistic (all others)

The graphical determination of trend-prediction ability (as defined by the matching of trend slopes) is an important consideration for any model. It would be impossible to predict with much greater accuracy since there are limitations in the availability of important input parameters, and since the system is an open domain. Since the actual and model trends are cumulative, it should be noted that larger errors between the actual and model predictions are also cumulative, thus explaining areas of relatively large discrepancy. But because the slopes during these periods are close to identical, they do represent a very good event-to-event predictive ability.

4.3 Event Prediction Models

Results for prediction of single events were not as favorable as anticipated. This was due to the step function of the breaks. With real data outputs which are stepped integrals, the ability of the model to predict a whole-number using a continuous value output proved to be exceedingly difficult, no matter which method of event selection criteria was chosen. The models, when presented this information, would assign a continuous value probability to the output. Use of a threshold output model would have been advantageous, however models provided in the *NeuroShell 2* program limited the output step to either 0 or 1, preventing the prediction of a multiple break, which is not the purpose of this study. However, one of the typical models shown in Figure 34 demonstrates that the model shows potential for future development.

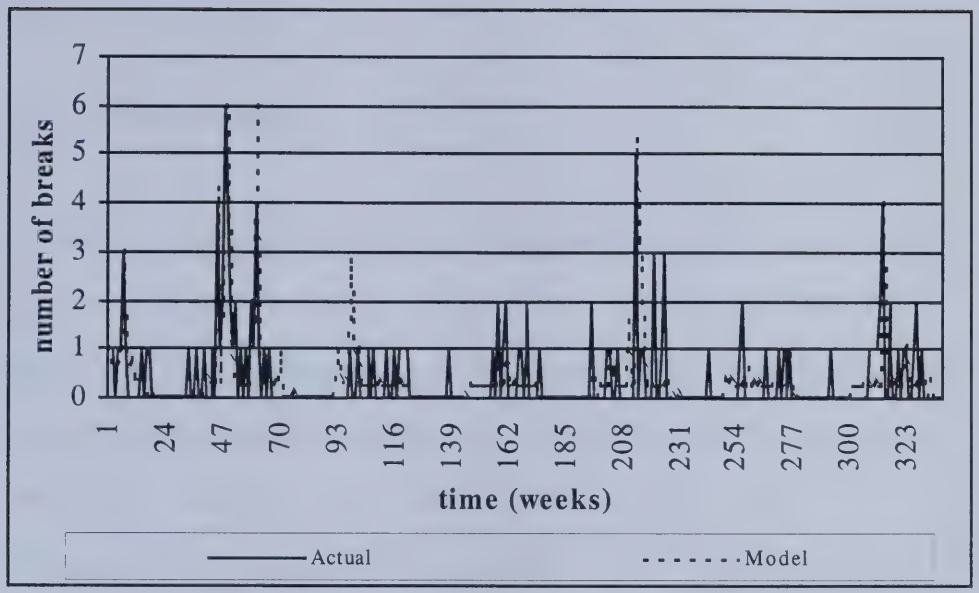


Figure 34. Typical event prediction results.

The models' single-event prediction capabilities were not well forecasted, using the R^2 criteria. Evaluation of these models using graphical analysis proved that the models could reasonably predict the probability of a break event, however the severity of the event was not necessarily matched. Given the purpose and level of research into this matter, initial results are reasonable. However, it is recommended that more research be conducted and more inputs be investigated.

5.0 DISCUSSION

5.1 Sensitivity Analysis

A sensitivity analysis was conducted on the chosen models to determine the robustness in the prediction of events that had not been previously presented to the model. The purpose of the sensitivity analysis was: to determine the model's actual learning ability; to determine the complexity of the models' learning pathways; and to confirm that the model extracted cause-effect logic underlying pipe failures, rather than pure memorization of the data. The following parameters were cumulatively tested for their robustness:

- 7-day average temperature (Model A and B);
- 7-day average temperature, lagged 1 week (Model A and B);
- 7 day water temperature (Model A and B);
- 7-day average (air-water) temperature differential (Model B only);
- 7-day average (air-water) temperature differential, lagged 1 week (Model A and B);
- 1-year previous historical break, moving total (Model A and B);
- 1-week previous historical break, moving total (Model B only);
- maximum 7-day temperature change (Model B only);
- maximum absolute 7-day temperature change (Model B only);
- maximum daily temperature change ; (Model B only)
- maximum daily temperature change, lagged 1 week (Model B only);

- maximum absolute daily temperature change (Model B only); and
- maximum absolute temperature change, lagged 1 week (Model B only).

Input parameters were varied in isolation from other independent input variables so that their effects on the models' output parameter could be identified for causality. Input parameters related to other input parameters were varied concurrently (e.g. 7-day average temperature and 7-day average temperature lagged 1 week were varied simultaneously) to maintain consistency of logic. Similarly, parameters related to the variable of interest were also adjusted, since varying only one parameter when it is related directly (or indirectly) to other variables would invariably "confuse" the model. For example, all temperature parameters (7-day average temperatures, 7-day maximum temperatures, etc.) were varied concurrently since they depend on the same raw air temperature data. Most of the input parameters were adjusted to 30 percent less than and 30 percent greater than the models' original inputted values. However, input parameters based on the number of breaks (previous week and previous year moving totals) were subtracted from the totals. Results of the sensitivity analysis are presented from Figure 35 through Figure 43.

Model A

Sensitivity analysis of air temperature parameters (varying 7-day average temperature and 7-day average temperature lagged one week) showed that a 30 percent increase in air temperature (both positive and negative magnitudes) resulted in a predicted percent increase of 25 percent from the actual values. A 30 percent decrease in air temperature

resulted in a decrease in cumulative pipe breaks of 19 percent, at the end of the 6 ½ year period. Slope changes were more pronounced during winter events for the 30 percent increases in air temperatures, conversely they were less pronounced for the 30 percent decreases.

The results from this analysis are logical. A percent increase in temperatures translates to increase in the magnitudes (i.e. the range of values is increased by the corresponding percent). It also translates to an increase in daily and weekly differences. This serves to magnify changes in temperature. Changes in temperature are significant for soil-pipeline interactions and frost heave action, which supports the previously discussed theories. Therefore the model tends to show a cause-effect relationship between temperature changes and pipe breaks.

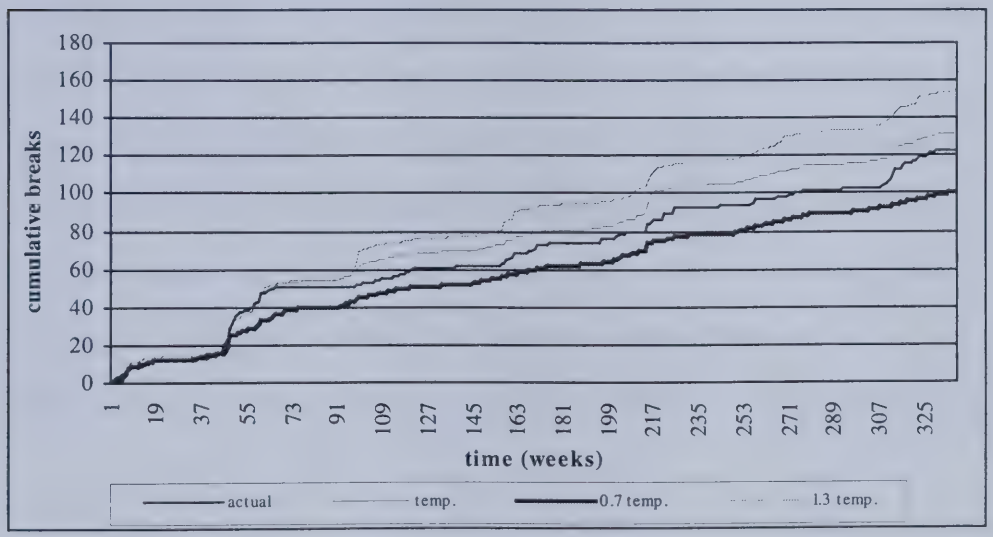


Figure 35. Model A sensitivity analysis: 7-day average air temperature.

Graphical analysis of the sensitivity of water temperatures for Model A (Figure 36) shows that a 30 percent decrease in water temperature results in a 44 percent increase in water main failures. This representation assumes that the air temperature range does not change during this manipulation, so that a smaller change in water temperature is not accompanied by a proportional change in air temperatures. This creates a differential temperature across the wall of the pipe (assuming air temperature gives a reasonable reflection of ground temperature). The resulting hoop stresses may result in longitudinal split and diagonal failures.

Analysis also shows that a 30 percent increase in water temperatures results in only a 2 percent increase in pipe failures. Possibly, the larger ranges in water temperatures correspond to the range of the air temperatures, creating a less significant pipe wall temperature differential. This assumption is logical since the same is assumed for the 30 percent decrease in water temperature. As mentioned, the 30 percent decrease would result in a smaller range of water temperature fluctuation. These smaller temperature changes (relative to the air temperature changes) again support the theory of a larger temperature differential. Another plausible reason for the reactions of this model to water temperature variations, is that Model A places more importance on the determination of pipe breaks (A reminder is that the model contains only 5 parameters).

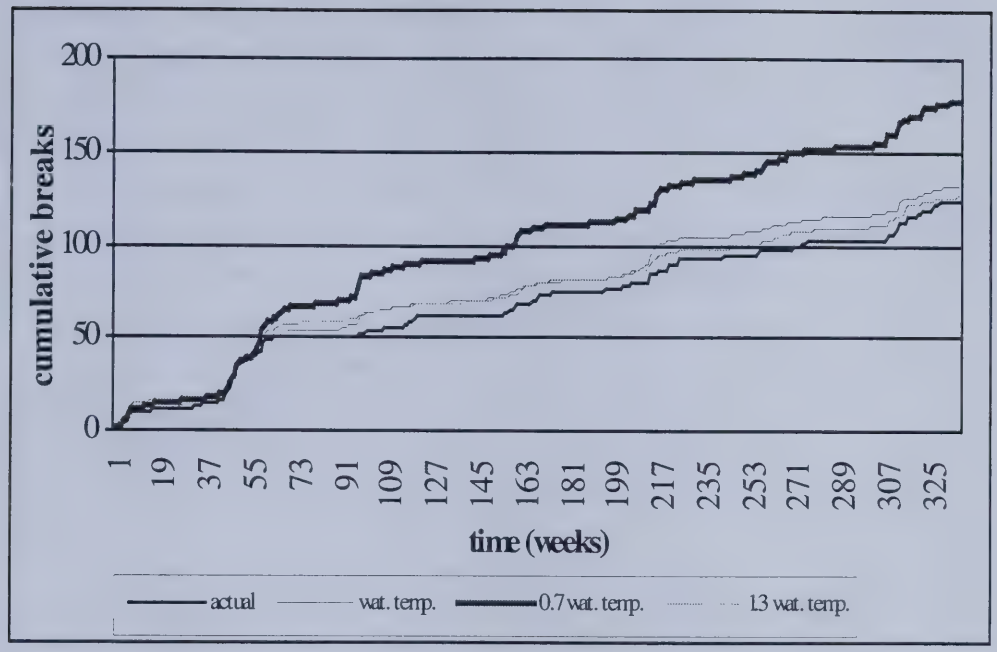


Figure 36. Model A sensitivity analysis: 7-day average water temperature.

Output effects of a 30 percent increase in pipe wall temperature differentials followed logical interpretations (Figure 37). As expected, a 30 percent increase in temperatures resulted in larger cumulative pipe breaks (a 26 percent increase from the actual value). This may be due to the exaggerated maximum and minimum temperatures, or it may be due to exaggerated temperature changes. Such changes would logically accelerate pipe failure mechanisms such as frost heave (having increased the rate of freezing) and soil-pipe interactions (long heated periods followed by cooler temperatures could indicate precipitation events, and therefore periods of soil instability). Conversely, a 30 percent temperature decrease would understate temperature ranges and changes, and thus indicate weaker pipe failure mechanics.

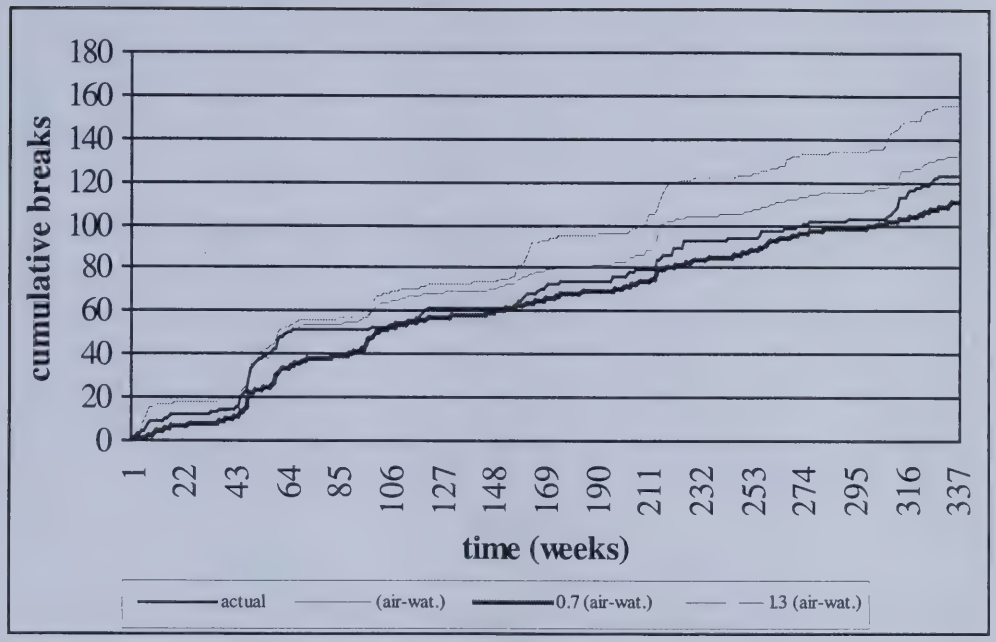


Figure 37. Model A sensitivity analysis: Pipe-wall temperature differential.

A lowered 1-year historical break total indicates that the ensuing break trend also decreases, as demonstrated by 16 percent drop in pipe failures. This is possibly due to greater structural integrity of the pipe system, a possible contributing factor being pipe wall corrosion. Results from a greater 1-year historical break history show a much more significant rise in the number of cumulative pipe breaks (49 percent increase). This model forecast is logical, indicating that structural integrity of the piping is an extremely important factor in pipe breaks.

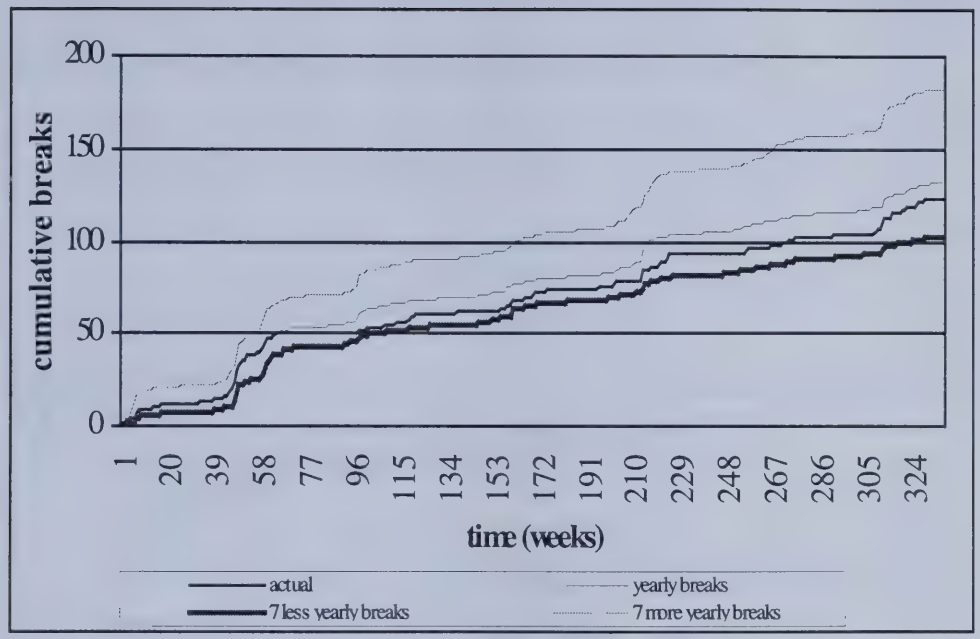


Figure 38. Model A sensitivity analysis: Historical 1-year break frequency.

Model B

A sensitivity analysis was also performed for Model B. Temperature was again analyzed, but more air temperature inputs are included with this model (Figure 39). Like Model A, Model B indicates that larger fluctuations in air temperature ranges and magnitudes result in extremely significant percent increases in pipe breaks. In fact, a 70 percent increase in pipe breaks is predicted by a 30 percent increase in air temperatures (over the 6 ½ years). Clearly, this model places great importance on the various air temperature parameters to indicate pipe failure mechanisms. Lowered ranges (30 percent decrease in air temperature range and magnitude) result in lesser cumulative pipe breaks

(13 percent decrease). As in Model A, this qualitatively supports the idea that frost heave and soil-pipeline interactions are major factors in water main failures.

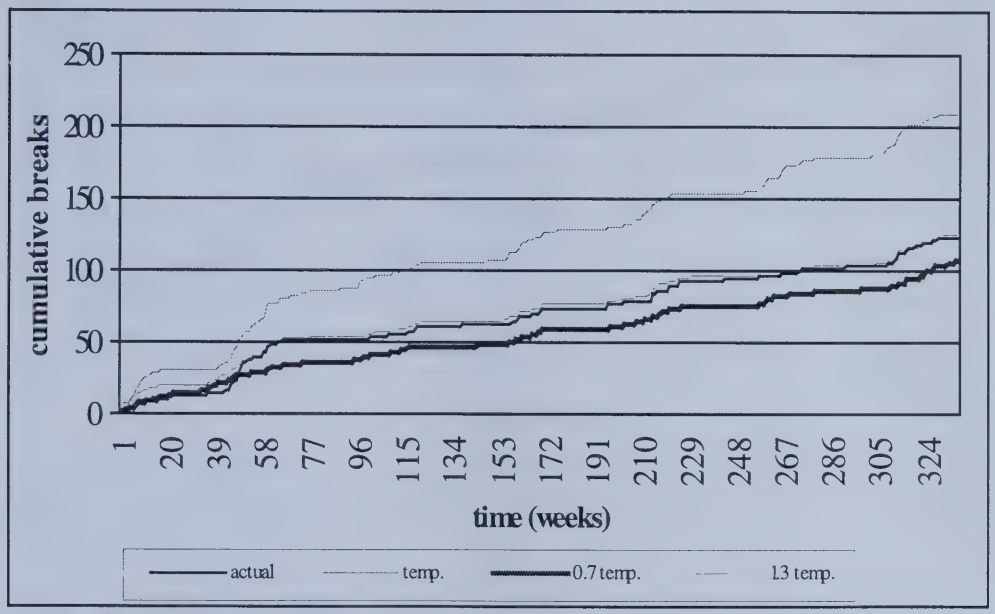


Figure 39. Model B sensitivity analysis: Air temperature.

Comparison of sensitivity analysis for water temperature (Figure 36 and Figure 40) shows that Model B does not place as much importance on water temperature. A 9 percent increase and 2 percent decrease are found for 30 percent increases and decreases, respectively, in water temperature. This does not agree qualitatively with the sensitivity analysis of Model A. It is apparent that Model B places less significance on the water temperature parameter.

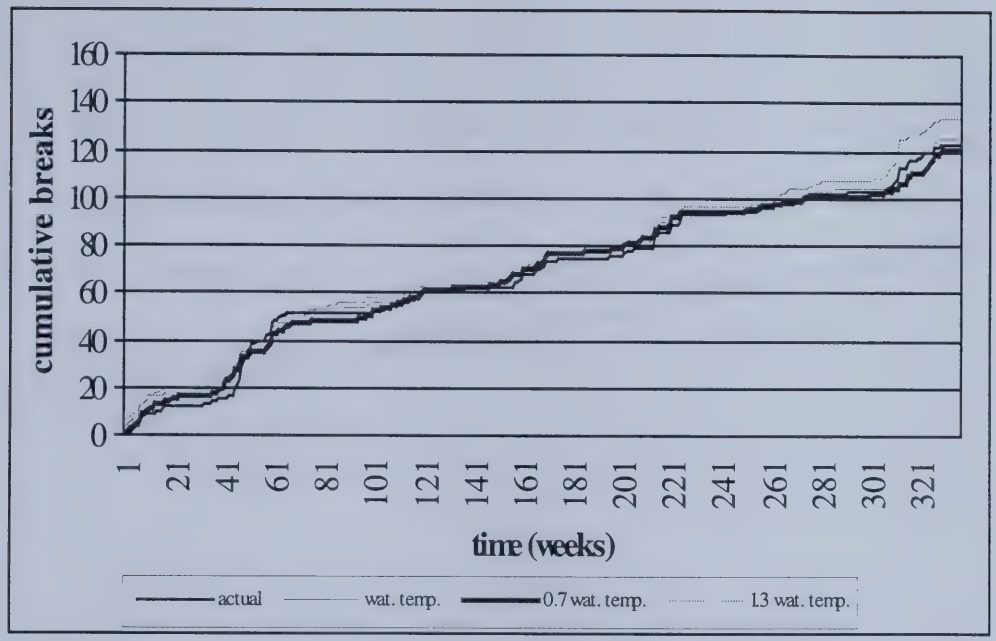


Figure 40. Model B sensitivity analysis: 7-day water temperature.

Results from the sensitivity analysis of pipe wall temperature differentials were extremely significant (see Figure 41). A 30 percent increase in the temperature differential results in a remarkable 91 percent increase in pipe break totals. A 30 percent decrease results in a 6 percent increase in pipe failures. For this, the trend of the breaks (steep slopes) indicates rapid numbers of failures during rapid temperature changes. One can infer that pipe wall temperature gradients are an important cause of pipe failures.

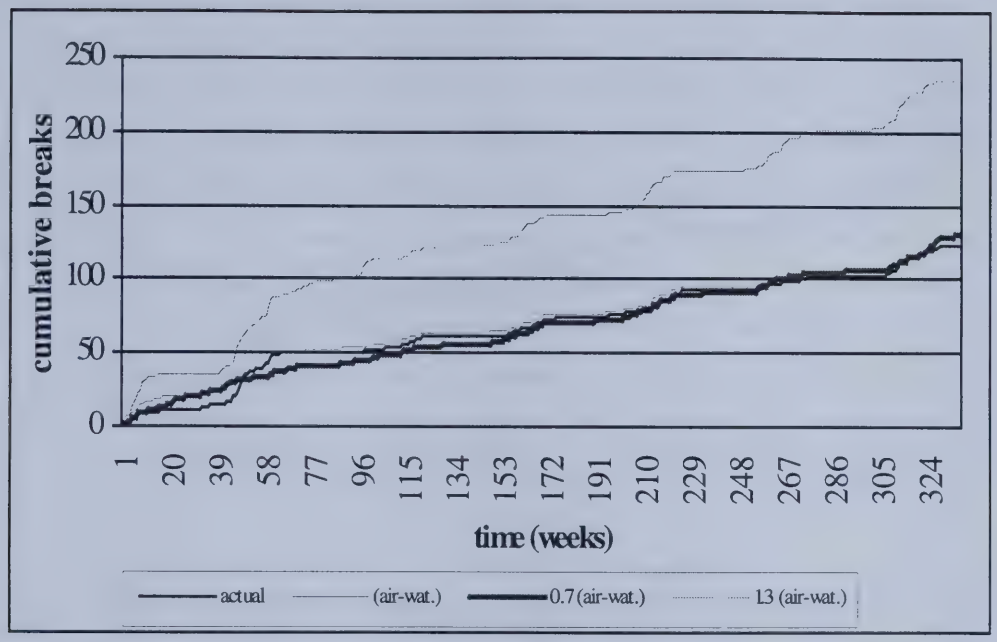


Figure 41. Model B sensitivity analysis: Pipe-wall temperature differential.

The previous week's pipe breaks (spatial cluster index) was formulated to try to mimic the phenomena of multiple pipe breaks. As literature indicates, results show that increases in breaks for the week previous increased the likelihood of pipe breaks in the present period (refer to Figure 42). In the same reasoning, no breaks in the week previous decrease the chances of subsequent multiple pipe failures occurring due to the spatial clustering. It may be inferred that multiple pipe failures for this area would decrease if the previous week had minimized pipe failures.

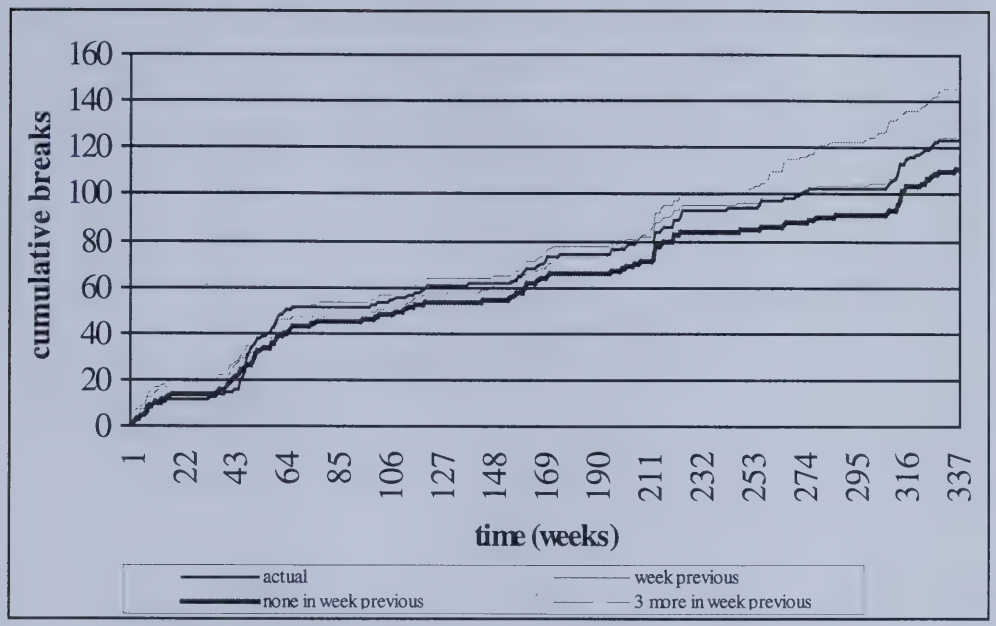


Figure 42. Model B sensitivity analysis: Spatial cluster index.

Finally, as in Model A, it was found that the previous year's total breaks was a good indication of the potential for future pipe failures occurring. A 7-break increase results in a 65 percent increase in pipe breaks. A 7-break decrease results in only a 6 percent decrease in water main failures. This large discrepancy can be explained in that pipe failures will occur inevitably, in large part due to temperature and corrosion effects. An uncharacteristically higher break total indicates that pipes are in a weakened state. Lower break totals indicate pipes are typically stronger, or perhaps it is indicative of work performed on the pipes (i.e., possible mitigation by cathodic protection, pipe section repair or replacement). The historical break total gives an indirect indication of the pipe's integrity, possibly inferring corrosion influences.

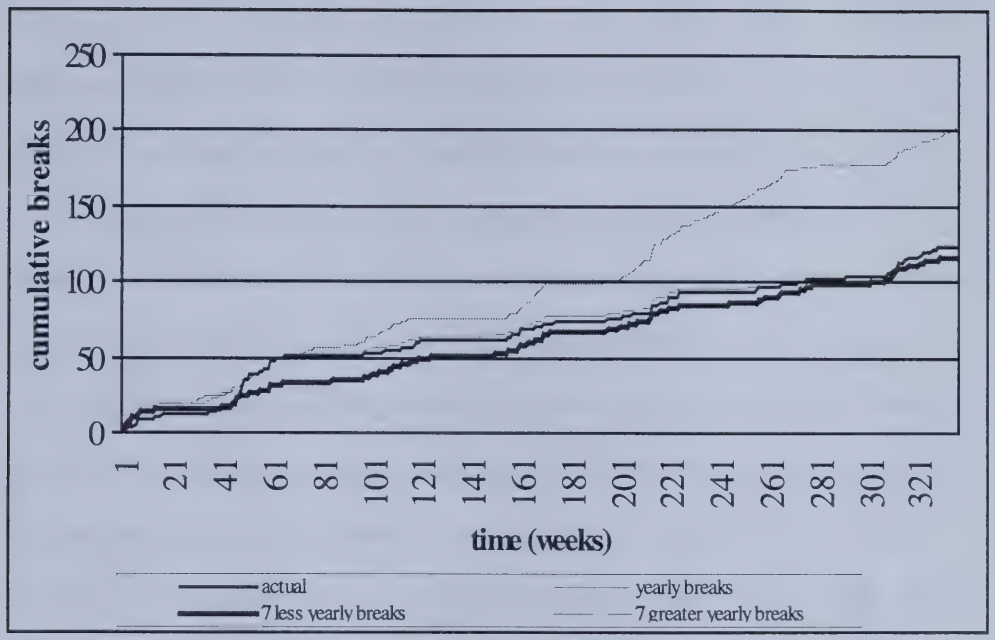


Figure 43. Model B sensitivity analysis: Historical 1-year break frequency.

5.2 Apparent Influential and Causal Factors

From the sensitivity analysis, there is evidence that the learned, intrinsic logic underlying the models developed with the ANN methodology is consistent. This is demonstrated by the relatively consistent trends predicted by the models, and by the sensitivity analyses' apparent support of the relationships between the input parameters and the pipe failure mechanisms. Effects of frost heave, soil-pipeline interaction and pipe wall temperature differentials are strongly related to the inputs presented. Inferences can be made from the 1-year previous historical breaks, spatial cluster index, and possibly for water temperature parameters.

Frost heave mechanisms are related to the rate of frost penetration. As no ground temperature data could be found, air temperature was used to characterize this phenomenon. As discussed earlier, this is a relatively good approximation if the information is given in time-series. By presenting the temperatures in time-order, the air temperature can indicate a rate of change, which the model can then intrinsically relate to ground temperature. A larger change in temperature and larger, negative magnitudes indicates that frost penetration rates will be higher. This leads to greater frost heaving, and therefore greater stresses applied to the pipe. Logically then, more pipe breaks should occur. Both models sensitivity analysis shows this (Figure 35 and Figure 39) to be the case. Conversely, smaller temperature changes and smaller temperature magnitudes have indicated less breaks.

Soil-pipeline interactions are caused by sudden drops in temperature. This type of behavior is consistent with the sensitivity analyses performed. Magnifying the air temperatures by a factor of 1.3 causes differences in temperature to be 1.3 times greater as well. This also translates to increases in both negative and positive magnitudes, and also magnification of changes in temperature. With this reasoning in mind, these magnifications should also result in more pipe failures. This has been demonstrated. In contrast, taking 0.7 (70 percent) of the values results in lower temperature changes and relatively smaller magnitudes. This results in less pipe failures, which is also illustrated in both models' analyses.

For these models, the pipe wall temperature gradients were depicted by the difference of the average air and average water temperatures. Air temperature is indicative of the ground temperature (which contacts the exterior of the pipe) and water temperature is indicative of the interior of the pipe. Differences in temperature result in higher hoop stresses exerted on the pipe. Increasing temperature differentials should increase stresses, resulting in more water main failures. Decreasing stresses should translate to lower stress levels, therefore, less failures. Both model A and model B sensitivity analyses (Figure 37 and Figure 42) reflect these effects on pipe break totals.

The historical 1-year previous break frequency indicates the study area's past year break history. It gives an indication of the stability of the system, therefore indicating the potential for future breaks. If less breaks occur in the previous year, this may be indicative of a sturdy pipe infrastructure, made durable by mitigative actions (i.e., pipe replacements, pipe repairs, or cathodic protection). More breaks may be indicative of increasing corrosion problems, or other activities that have caused pipe instability. Therefore, increasing the past year's break total indicates more instability. Results from both models indicate these trends are consistent (Figure 38 and Figure 43).

Analyses of the spatial cluster index (1-week previous pipe break total) indicates a logic similar to that of the historical 1-year break frequency. Decreasing the number of breaks to zero breaks indicates lower breaks in the upcoming week. Increasing the number of breaks by three breaks shows an increase in break frequency in the upcoming week. This clustering phenomenon was attributed to disturbances in the surrounding soils, causing

further instability and settlement. This instability increases the likelihood of pipe failures. Although this relationship can only be inferred, it seems to be a plausible justification.

Results from the sensitivity analysis of the water temperature parameter do not yield completely consistent results. While both models (Figure 36 and Figure 40) demonstrate that increased water temperatures (ranges and magnitudes) show negligible effects on pipe breaks, decreases in water temperatures show a significant increase in pipe breaks for Model A only. For Model B, significance of lowered water temperature ranges and magnitudes is also negligible.

In order to rationalize the logic of the model, it must be accounted that water temperatures will always be positive values (since water freezes at $0^{o}C$). Magnifying values by a factor of 1.3 depicts warmer water temperatures. Magnifying by a factor of 0.7 results in a smaller temperature range, and cooler year-round water temperatures. Therefore, values magnified by a factor of 1.3 are interpreted as warmer temperatures year-round, which is less conducive to pipe breaks. Values magnified by a factor of 0.7 translate to cooler water temperatures year-round, which is more conducive to pipe failures. Since Model A has only 5 parameters, compared to Model B's 13 parameters, Model A must infer more information from less parameters, which may be somewhat simplistic. The fact Model B has more than double the inputs of Model A also explains why the percent change of cumulative pipe breaks (for each common input parameter) is

different. However, qualitative analysis is in agreement, which is the primary concern for this modeling study.

From the models sensitivity analyses presented above, one can infer significant findings. The models analyses support the literature citing the nature of pipe break mechanics, and they generally adhere to the described failure mechanisms. Therefore, significance of the different input parameters is generally understood.

5.3 Model Capabilities, Limitations, and Uses

Quantitative and qualitative analysis of the two models indicate evidence that the Artificial Neural Networks methodology is capable of predicting water main failures. The R^2 statistics and trend predicting abilities of the models indicate the potential for developing accurate, forecast-capable models. The sensitivity analysis shows that intrinsic logic of the various failure mechanisms is credibly captured. Manipulation of the model input parameters also allows for inferences to the prevailing failure mechanism tendencies of the study area. Having done this, it is possible to implement the appropriate mitigative techniques.

Due to assumptions made during source data analysis and data collection, these models are most applicable to specific situations that are directly related to the reliability and availability of raw source data. The assumption that the area was composed of a uniform soil type does not allow for this model to be applied to non-uniform soils, nor is it

necessarily applicable to different soil types. Since modeling was restricted to the 150 mm diameter cast iron pipe, modeling of larger-size pipes, or different pipe materials (i.e., asbestos cement, PVC, etc.) may require extensive modification of input parameters because different failure mechanisms will dominate. For the same reasons as the above assumptions, application to warmer climates may warrant use of different type of parameters (to reflect the dominance of failure mechanisms for warmer climates).

Further to this, these models are limited by the open system nature of the study. The model will not predict random events that cause pipe breaks such as water hammer events and severe weather events. However, it may be possible to determine what percentage of all pipe breaks will be of a random nature, and use this information as a prediction tolerance. Further study is required to perform this.

These models are dependent on the historical information of the study area, and therefore is specific to the area studied. These models must be retrained for other subdivisions. The model results presented do favorably indicate the application of the model for the study of pipe breaks. However, the models themselves cannot be viewed as applicable tools to the entire City of Edmonton's cast iron water distribution system.

The models developed illustrate the utility of using Artificial Neural Network methodology for predicting pipe breaks. While the model is limited in application for the above reasons, these models, along with newly developed ANN models, have valuable uses. Development of ANN pipe break models, for different subdivisions within

Edmonton, will make it possible to prioritize areas for the Cast Iron Renewal Program. The model itself may be useful as a monitoring tool, to evaluate the progress of past water main renewal efforts. As mentioned earlier, manipulation of model input parameters would allow inferences as to pipe failure causes, and to allow for appropriate mitigation techniques.

Potential use of the model for pipe break prediction must also include a methodology to forecast the input parameters, namely air and water temperatures. The models use a full array of accessible information to accurately forecast probabilities of pipe breaks in time series. Therefore, input parameters should also be able to indicate severe or abnormal conditions. This will permit for more accurate forecasts of pipe breaks. Once this can be accomplished, implementation of the model for practical purposes becomes possible.

6.0 CONCLUSIONS AND RECOMMENDATIONS

With respect to developing a prediction-capable pipe break model for a given city subdivision, the goal has been achieved. Utilizing readily available information, the Artificial Neural Network methodology has proven its feasibility in this regard. The forecasting ability of this model has been demonstrated using the Calder subdivision as a model study. Quantitative analysis using R^2 statistics and visual examination of trends (slope matching) has permitted appropriate selection of models based on accuracy and trends matching capability. Qualitative analysis, in the form of sensitivity analysis, was performed to demonstrate the ANN models' ability to "learn" the intrinsic logic underlying pipe failure mechanisms. Through the careful model development methodology and evaluation, coupled with the sensitivity analysis, it is shown that concepts are being extracted from the input parameters, rather than the model purely "memorizing" the inputted data in order to predict the output.

The ANN models also demonstrate their potential applicability as screening tools. The developed models were able to accurately predict the cumulative number of pipe failures for the six and a half year study time period. Application to other city subdivisions would offer comparative information for priority setting for the Cast Iron Renewal Program.

Manipulation of input parameters in developed models permit its use for inferring which of the cast iron pipe failure mechanics dominate. From the sensitivity analysis results, it

is clear that frost heave, soil-pipeline interaction and pipe wall temperature gradients are responsible for a large portion of the 150 mm cast iron water main failures. Therefore, air temperature plays a very significant role in the model prediction. Other parameters show significance, but further investigation is required to make quantifiable conclusions.

Since the developed models do not include a number of specific measures thought to be important to pipe failures, the models developed in this exercise are not complete. While they do demonstrate the utility of using Artificial Neural Networks for predicting pipe breaks, further work for data collection and model development is required to ensure the model is flexible for future applications. From the perspective of frost action, considerations may include further examination of the freezing index as potential model input parameter. This index would provide a generalization of the severity of a winter event. Another possibility is to examine the effect of pipe-trench backfill material and ground surface (e.g., asphalt, clay cover, gravel cover, or other). Both backfill materials and ground surface will change the thermal exchange of heat from ground to air, thus varying the effect of temperature transitions on pipe breaks.

Having made the above conclusions, it is clear that more work is required to facilitate future use. However, the models presented have useful applications for the Calder area. Given that both Models A and B demonstrated exceptional accuracy in prediction, these models may be used to diagnose existing problems in the area. These ANN models are capable of predicting the frequency of pipe failures caused by frost heave, soil-pipeline interaction and pipe-wall temperature gradients. If there is an uncharacteristically-large

discrepancy between the actual number of pipe breaks and the number of pipe breaks predicted by the ANN models, it is apparent that there are other failure mechanisms contributing to pipe breaks in the area. This may include corrosion problems or operating pressure-related problems (i.e., water hammer events or pumping problems), or circumstance requiring detailed investigations. In any case, the model will be a useful information tool for diagnosing this possibility.

This model study illustrates the need for the following actions, to facilitate ease, and more comprehensive development of Artificial Neural Network models for water main failure prediction:

1. Inclusion of more descriptive data;
2. Collection technique improvements of present data;
3. Characterization of temperature data and;
4. Exploration of input parameter importance, and special phenomena.

To further develop ANN models that are accurate and flexible, inclusion of more descriptive data is needed. Initial models required making assumptions that were scope-limiting since it required constant values. The availability of detailed soils parameters, physical pipe characteristics, and in-situ pipe conditions would be assets. Soils parameters and physical pipe characteristics would indicate more explicitly the characteristics of the cast iron pipe. In-situ pipe conditions, possibly collected from hydroscope measurements, may also be of value. Overall, the goal of inclusion of these

parameters would be to widen the scope of application of the models, instead of limiting areas of application.

Source data collection in this study demonstrated a need for more complete and accurate data. Much of the data used for this study was available only in hardcopy and it was difficult to obtain, or was of insufficient detail or quantity. As a large amount of quality data is required for ANN application, it is recommended that more complete databases are kept, and this information is updated, to facilitate ease of collection of raw data.

Due to the importance of air temperature to the ANN models developed, it is recommended that weather data be characterized such that typical years, above- and below-average temperatures and other special events be characterized, and therefore used for pipe break sensitivity analysis. Inclusion of a freezing index may be a possible avenue for characterizing weather. The sensitivity analysis tool then potentially becomes more valuable when pipe break rates between areas are similar.

To accurately quantify the effect of certain input parameters for a given area, it will be necessary to develop the ANN models, using the above study as an initial starting base. As demonstrated in this study, the feasibility of Artificial Neural Networks methodology was proven, and is effective as a diagnosis tool. However, subsequent models with more descriptive parameters will enhance the understanding of the effects of individual causal or influencing input parameter on cast iron water main failures.

REFERENCES

Anderson, D.M. and Tice, A.R., 1972. Predicting Unfrozen Water Contents in Frozen Soils From Surface Area Measurements. Highway Research Record, 393: pp. 12-18.

Anderson, D.M., Williams, P.J., Guymon, G.L. and Kane, D.L., 1984. Principles of Soil Freezing and Frost Heaving. Frost Action and Its Control. American Society of Civil Engineers, Hanover, New Hampshire, pp. 1-15 pp.

Bahmanyar, G. and Edil, T.B., 1983. Cold Weather Effects on Underground Pipeline Failures. In: M.B. Pickell (Editor), Pipelines in Adverse Enviroments II. American Society of Civil Engineers, San Diego, California, pp. pp. 579-593.

Bates, J., Cassaro, M.J. and Cassaro, M.A., 1996. Damage Analysis of a Water Distribution System Subject to Natural Hazards. In: G.W. Housner and R.M. Chung (Editors), Natural Disaster Reduction. American Society of Civil Engineers, Washington, D.C., pp. pp. 357-358.

Booth, G.H., Cooper, A.W., Cooper, P.M. and Wakerley, D.S., 1967. Criteria of Soil Aggressiveness Towards Buried Metals. British Corrosion Journal, 2(May): pp. 104-118.

Boyd, D.W., 1973. Normal Freezing and Thawing Degree-Days for Canada 1931-1960. Normal Freezing and Thawing Degree-Days for Canada 1931-1960: pp. 1-3.

Burrows, R. and Qiu, D.Q., 1995. Effect of Air Pockets on Pipeline Surge Pressure. Proceedings of the Institution of Civil Engineers: Water, Maritime and Energy, 112(Decembert): pp. 349-361.

Clark, C.M., 1971. Expansive-Soil Effect on Buried Pipe. American Water Works Association Journal, 63(July): pp. 424-427.

Dorn, R.A., 1989. Conducting a Waterline Corrosion Evaluation. In: J.F. Malina Jr. (Editor), National Conference on Environmental Engineering. American Society of Civil Engineers, Austin, Texas, pp. pp. 168-175.

Garrett, J.H.J., Ghaboussi, J. and Wu, X., 1992. Neural Networks. Expert Systems for Civil Engineers: Knowledge Representation. American Society of Civil Engineering, New York, New York.

Goulter, I., Davidson, J. and Jacobs, P., 1990. Predicting Watermain Breakage Rates. In: J.F. Scott and R.M. Khanbilvardi (Editors), Water Resources Infrastructure: Needs, Economics, and Financing. American Society of Civil Engineers, Fort Worth, Texas, pp. pp. 66-69.

Goulter, I.C. and Kazemi, A., 1988. Spatial and Temporal Groupings of Water Main Pipe Breakage in Winnipeg. Canadian Journal of Civil Engineering, 15: pp. 91-97.

Goulter, I.C. and Kazemi, A., 1989. Analysis of Water Distribution Pipe Failure Types in Winnipeg, Canada. Journal Transportation Engineering, 115(2)(March): pp. 95-111.

Habibian, A., 1994. Effect of Temperature Changes on Water-Main Breaks. Journal of Transportation Engineering, 120(2)(March/April): pp. 312-321.

Hu, J. and Selvadurai, A.P.S., 1995. Influence of Tertiary Creep on the Uplift Behaviour of a Pipe Embedded in a Frozen Soil. In: J.K. Jeyapalan and M. Jeyapalan (Editors), Advances in Underground Pipeline Engineering. American Society of Civil Engineers, Bellevue, Washington, pp. pp. 345-358.

Jarvis, M.G. and Hedges, M.R., 1994. Use of Soil Maps to Predict the Incidence of Corrosion and the Need for Iron Mains Renewal. Water and Environmental Management, 8(1)(February): pp. 68-75.

Jones, R.H., 1995. Properties of Freezing, Frozen and Thawed Soils. In: J.S. Harris (Editor), Ground Freezing in Practice. Thomas Telford, New York, New York, pp. pp. 26-69.

Kathol, C.P. and McPherson, R.A., 1975. Urban Geology of Edmonton 1975. Bulletin 32, Edmonton, Alberta.

Kettler, A.J. and Goulter, I.C., 1985. An Analysis of Pipe Breakage in Urban Water Distribution Networks. Canadian Journal of Civil Engineering, 12: pp. 286-293.

Kitaura, M. and Miyajima, M., 1996. Damage to Water Supply Pipelines. Soils and Foundations(January): pp. 325-333.

Konrad, J.-M., 1987. Procedure for Determining the Segregation Potential of Freezing Soils. Geotechnical Testing Journal, 10(2)(June): pp. 51-58.

Konrad, J.-M., 1994. Sixteenth Canadian Geotechnical Colloquium: Frost Heave in Soils: Concepts and Engineering. Canadian Geotechnical Journal, 31: pp. 223-245.

Konrad, J.-M. and Morgenstern, N.R., 1980. A Mechanistic Theory of Ice Lens Formation in Fine-Grained Soils. Canadian Geotechnical Journal, 17: pp. 473-486.

Konrad, J.-M. and Morgenstern, N.R., 1981. The Segregation Potential of a Freezing Soil. Canadian Geotechnical Journal, 18: pp. 482-491.

Konrad, J.-M. and Nixon, J.F., 1994. Frost Heave Characteristics of a Clayey Silt Subjected to Small Temperature Gradients. Cold Regions Science and Technology, 22: pp. 299-310.

Kujala, K., 1993. Evaluation of Factors Affecting Frost Susceptibility in Soils. In: A. Phukan (Editor), Frost in Geotechnical Engineering. A.A. Balkema, Anchorage, Alaska, pp. pp. 83-87.

Kujala, K. and Laurinen, K., 1989. Freeze-Thaw Effects on Thaw Settlement and Pore Pressure. In: H. Rathmayer (Editor), Frost in Geotechnical Engineering. Technical Research Centre of Finland, Saariselka, Finland, pp. pp. 523-533.

Kurilko, A.S., Kravtsova, O.N., Stepanov, A.V. and Timofeez, A.M., 1989. The Effect of Freezing-Thawing Cycles on Soil Mass Transfer and Thermal Characteristics. In: H. Rathmayer (Editor), Frost in Goetechnical Engineering. Technical Research Centre of Finland, Saariselka, Finland, pp. pp. 311-321.

Lutey, R.W. and Mason, P.D., 1994. Report: Identification of Root Cause Failure op Piping in a Service Water System, International Joint Power Generation Conference. American Society of Mechanical Engineers, pp. pp. 69-78.

McGaw, R., 1972. Frost Heaving Versus Depth to Water Table. Highway Research Record, 393: pp. 45-55.

Miller, R.D., 1972. Freezing and Heaving of Saturated and Unsaturated Soils. Highway Research Record, 393: pp. 1-11.

Milligan, G., 1995. Practical Examinations. Ground Engineering, 28(December): pp. 19-20.

Moncarz, P.D., Shyne, J.C. and Derbalian, G.K., 1987. Failures of 108-Inch Steel Pipe Water Main. Journal of Performance of Constructed Facilities, 1(3)(August): pp. 168-187.

Morris Jr., R.E., 1967. Principal Causes and Remedies of Water Main Breaks. Journal American Water Works Association, 59(July): pp. 782-798.

Nixon, J.F., 1994. Role of Heave Pressure Dependency and Soil Creep in Stress Analysis for Pipeline Frost Heave. In: D.W. Smith and D.C. Sego (Editors), Cold Regions Engineering. Canadian Society for Civil Engineers, Edmonton, Alberta, pp. pp. 397-412.

O'Day, K., 1982. Organizing and Analyzing Leak and Break Data for Making Main Replacement Decisions. American Water Works Association Journal, 74(11)(Novembert): pp. 589-594.

O'Farrell, S.J., 1995. Water Main Failure, Renewal and Cathodic Protection Issues in the City of Edmonton. CLDREP95.DOC, Public Works Department

Pawluk, S., 1988. Freeze-Thaw Effects on Granular Structure Reorganization for Soil Materials of Varying Texture and Moisture Content. Canadian Journal of Soil Science, 68: pp. 485-494.

Penner, E., 1972. Influence of Freezing Rate on Frost Heaving. Highway Research Record, 393: pp. 56-64.

Phukan, A., 1985. Frozen Ground Engineering. Prentice-Hall International Series in Civil Engineering and Engineering Mechanics. Prentice-Hall, Englewood Cliffs, New Jersey, 336 pp.

Quraishi, A.A. and Al-Amry, M.S., 1992. Transportation of Demineralized Water: Case Study. Journal of Transportation Engineering, 118 (4)(July/August): pp. 576-585.

Rajani, B., 1992. Deformation of Pipelines in Frozen Soil. Ph.D. Thesis, University of Alberta, Edmonton, Alberta.

Rajani, B., Zhan, C. and Kuraoka, S., 1996. Pipe-soil interaction analysis of jointed water mains. Canadian Geotechnical Journal, 33: pp.393-404.

Rajani, B.B., Robertson, P.K. and Morgenstern, N.R., 1995. Simplified Design Methods for Pipelines Subject to Transverse and Longitudinal Soil Movements. Canadian Geotechnical Journal, 32: pp. 309-323.

Roy, M., Crispin, J., Konrad, J.M. and Larose, G., 1992. Field Study of Two Road Sections During a Freeze-Thaw Cycle. Transportation Research Record, 1362: pp. 71-78.

Sacluti, F., Stanley, S.J. and Zhang, Q., 1998. Use of Artificial Neural Networks to Predict Water Distribution Pipe Breaks, Western Canada Water and Wastewater Association, 50th Annual Conference. Western Canada Water and Wastewater Association, Calgary, AB.

Schmuller, J., 1990. Neural Networks and Environmental Applications. In: J.M. Hushon (Editor), Systems for Environmental Applications, ACS Symposium Series 431. American Chemical Society, Washington, D.C., pp. 235.

Selig, E.T., 1988. Soil Parameters for Design of Buried Pipeline. In: B.A. Bennett (Editor), Pipeline Infrastructure. American Society of Civil Engineers, Boston, Massachusetts, pp. pp. 99-116.

Shen, M. and Ladanyi, B., 1991. Soil-Pipe Interaction During Frost Heaving Around a Buried Chilled Pipeline. In: D.S. Sodhi (Editor), Cold Regions Engineering. American Society of Civil Engineers, West Lebanon, NH, pp. pp. 11-21.

Smith, W.H., 1968. Soil Evaluation in Relation to Cast-Iron Pipe. American Water Works Association Journal, 60(2)(February): pp. 221-227.

Ward Systems Group, I., 1993. NeuroShell 2 User's Manual. Ward Systems Group, Inc., Frederick, MD.

Yen, B.C., Tsao, C. and Hinkle, R.D., 1981. Soil-Pipe Interaction of Heated Pipelines. Transportation Engineering Journal(January): pp. 1-14.

Zhang, Q., 1996. Artificial Neural Networks and Their Application To Water Treatment. M.Sc. Thesis, University of Alberta, Edmonton, Alberta.

APPENDIX A. Sample Model Input Data, Model A.

Date	Temp,7d Average	Temp,n-7d 1 week lag	wat temp,7d Average	T7(a-w),n-7d 1 week lag	150,n-1yr 1 yr prev.
01/07/85	-2.3	-26.3	9.7	-35.8	15
01/14/85	-8.8	-2.3	9.7	-12.0	15
01/21/85	-9.9	-8.8	9.7	-18.5	16
01/28/85	-4.8	-9.9	9.7	-19.6	16
02/04/85	-23.4	-4.8	9.8	-14.5	17
02/11/85	-23.0	-23.4	9.9	-33.2	18
02/18/85	-6.2	-23.0	9.9	-32.9	21
02/25/85	-3.8	-6.2	9.9	-16.1	22
03/04/85	-4.7	-3.8	10.0	-13.7	22
03/11/85	-3.1	-4.7	10.1	-14.7	21
03/18/85	3.5	-3.1	10.1	-13.2	21
03/25/85	1.3	3.5	10.1	-6.6	21
04/01/85	2.5	1.3	10.1	-8.8	21
04/08/85	3.4	2.5	10.1	-7.6	21
04/15/85	7.6	3.4	10.1	-6.7	21
04/22/85	4.1	7.6	10.1	-2.5	22
04/29/85	4.7	4.1	10.1	-6.0	23
05/06/85	12.1	4.7	15.1	-5.4	23
05/13/85	11.2	12.1	15.1	-3.0	23
05/20/85	15.5	11.2	15.1	-3.9	23
05/27/85	14.4	15.5	15.1	0.4	23
06/03/85	11.1	14.4	15.8	-0.7	23
06/10/85	13.4	11.1	16.7	-4.7	23
06/17/85	13.7	13.4	16.7	-3.3	22
06/24/85	14.3	13.7	16.7	-3.0	22
07/01/85	16.7	14.3	16.7	-2.4	22
07/08/85	21.6	16.7	21.9	0.0	22
07/15/85	18.8	21.6	21.9	-0.3	22
07/22/85	17.9	18.8	21.9	-3.1	22
07/29/85	16.7	17.9	21.9	-4.0	22
08/05/85	21.0	16.7	19.3	-5.2	22
08/12/85	13.0	21.0	18.3	1.7	21
08/19/85	15.2	13.0	18.3	-5.3	22
08/26/85	14.5	15.2	18.3	-3.1	22
09/02/85	13.2	14.5	16.3	-3.8	22
09/09/85	7.7	13.2	11.2	-3.1	23
09/16/85	12.2	7.7	11.2	-3.5	23
09/23/85	4.7	12.2	11.2	1.0	23
09/30/85	5.6	4.7	11.2	-6.5	24
10/07/85	5.6	5.6	6.0	-5.6	24
10/14/85	4.1	5.6	6.0	-0.4	24
10/21/85	6.4	4.1	6.0	-1.9	23
10/28/85	3.9	6.4	6.0	0.4	24
11/04/85	0.8	3.9	5.9	-2.1	24
11/11/85	-10.3	0.8	5.9	-5.2	27
11/18/85	-6.9	-10.3	5.9	-16.2	27
11/25/85	-22.4	-6.9	5.9	-12.8	29
12/02/85	-21.8	-22.4	5.8	-28.3	34
12/09/85	-7.5	-21.8	5.6	-27.6	38
12/16/85	-8.3	-7.5	5.6	-13.1	40

Date	Temp,7d Average	Temp,n-7d 1 week lag	wat temp,7d Average	T7(a-w),n-7d 1 week lag	150,n-1yr 1 yr prev.
12/23/85	2.8	-8.3	5.6	-13.9	39
12/30/85	-2.2	2.8	5.6	-2.8	40
01/06/86	-7.7	-2.2	5.5	-7.8	37
01/13/86	-0.1	-7.7	5.5	-13.2	38
01/20/86	-3.2	-0.1	5.5	-5.6	36
01/27/86	-7.1	-3.2	5.5	-8.7	37
02/03/86	-6.7	-7.1	5.5	-12.6	36
02/10/86	-11.9	-6.7	5.4	-12.2	37
02/17/86	-16.6	-11.9	5.4	-17.3	35
02/24/86	-17.7	-16.6	5.4	-22.0	38
03/03/86	5.0	-17.7	5.1	-23.1	39
03/10/86	-3.8	5.0	4.8	-0.1	39
03/17/86	0.7	-3.8	4.8	-8.6	40
03/24/86	2.8	0.7	4.8	-4.1	41
03/31/86	5.4	2.8	4.8	-2.0	41
04/07/86	6.6	5.4	5.1	0.6	41
04/14/86	-1.1	6.6	5.1	1.5	41
04/21/86	7.8	-1.1	5.1	-6.2	40
04/28/86	6.1	7.8	5.1	2.7	40
05/05/86	5.4	6.1	10.4	1.0	39
05/12/86	9.2	5.4	12.5	-5.0	39
05/19/86	10.2	9.2	12.5	-3.3	39
05/26/86	16.1	10.2	12.5	-2.3	39
06/02/86	21.8	16.1	14.0	3.6	39
06/09/86	15.9	21.8	17.9	7.7	39
06/16/86	13.9	15.9	17.9	-2.0	39
06/23/86	17.4	13.9	17.9	-4.0	39
06/30/86	15.5	17.4	17.9	-0.5	39
07/07/86	15.8	15.5	16.9	-2.4	39
07/14/86	15.9	15.8	16.9	-1.1	39
07/21/86	16.4	15.9	16.9	-1.0	39
07/28/86	16.0	16.4	16.9	-0.5	39
08/04/86	17.5	16.0	18.3	-0.9	39
08/11/86	17.5	17.5	19.3	-0.7	39
08/18/86	17.4	17.5	19.3	-1.8	38
08/25/86	14.5	17.4	19.3	-1.9	38
09/01/86	17.2	14.5	19.3	-4.8	38
09/08/86	11.1	17.2	12.2	-2.1	38
09/15/86	7.4	11.1	12.2	-1.1	37
09/22/86	8.3	7.4	12.2	-4.8	37
09/29/86	8.7	8.3	12.2	-3.9	37
10/06/86	7.5	8.7	7.8	-3.5	36
10/13/86	6.9	7.5	7.8	-0.3	36
10/20/86	10.9	6.9	7.8	-0.9	36
10/27/86	9.3	10.9	7.8	3.1	35
11/03/86	0.1	9.3	6.5	1.5	35
11/10/86	-7.7	0.1	4.8	-6.4	34
11/17/86	-13.6	-7.7	4.8	-12.5	32
11/24/86	-10.3	-13.6	4.8	-18.4	31
12/01/86	-4.0	-10.3	4.8	-15.1	26

Date	Temp,7d Average	Temp,n-7d 1 week lag	wat temp,7d Average	T7(a-w),n-7d 1 week lag	150,n-1yr 1 yr prev.
12/08/86	-8.2	-4.0	5.1	-8.8	22
12/15/86	-3.8	-8.2	5.1	-13.3	19
12/22/86	-3.8	-3.8	5.1	-8.9	17
12/29/86	-1.9	-3.8	5.1	-8.9	16
01/05/87	-3.8	-1.9	5.2	-7.0	15
01/12/87	-2.8	-3.8	5.3	-9.1	15
01/19/87	-6.1	-2.8	5.3	-8.1	15
01/26/87	-5.2	-6.1	5.3	-11.4	15
02/02/87	-4.6	-5.2	5.3	-10.5	15
02/09/87	-0.1	-4.6	5.4	-10.0	13
02/16/87	-3.1	-0.1	5.4	-5.5	12
02/23/87	-1.1	-3.1	5.4	-8.5	9
03/02/87	-10.3	-1.1	5.4	-6.5	8
03/09/87	-7.5	-10.3	5.3	-15.7	8
03/16/87	-6.6	-7.5	5.3	-12.8	7
03/23/87	-1.6	-6.6	5.3	-11.9	8
03/30/87	-2.1	-1.6	5.3	-6.9	7
04/06/87	6.7	-2.1	3.9	-7.4	7
04/13/87	4.7	6.7	3.9	2.8	8
04/20/87	6.8	4.7	3.9	0.8	9
04/27/87	9.4	6.8	3.9	2.9	10
05/04/87	14.2	9.4	9.8	5.5	10
05/11/87	14.4	14.2	14.3	4.3	10
05/18/87	11.3	14.4	14.3	0.1	10
05/25/87	9.2	11.3	14.3	-3.0	10
06/01/87	14.7	9.2	14.3	-5.1	10
06/08/87	15.7	14.7	18.1	0.4	10
06/15/87	18.2	15.7	18.1	-2.4	10
06/22/87	17.5	18.2	18.1	0.1	10
06/29/87	16.8	17.5	18.1	-0.6	10
07/06/87	16.8	16.8	18.7	-1.3	10
07/13/87	15.9	16.8	18.7	-1.9	10
07/20/87	14.7	15.9	18.7	-2.8	10
07/27/87	19.5	14.7	18.7	-4.0	10
08/03/87	18.5	19.5	17.7	0.8	10
08/10/87	16.2	18.5	16.3	0.9	10
08/17/87	11.9	16.2	16.3	-0.1	10
08/24/87	13.1	11.9	16.3	-4.4	11
08/31/87	13.9	13.1	16.3	-3.2	11
09/07/87	14.3	13.9	15.2	-2.4	11
09/14/87	14.7	14.3	15.2	-0.9	11
09/21/87	12.8	14.7	15.2	-0.5	11
09/28/87	13.0	12.8	15.2	-2.4	11
10/05/87	14.0	13.0	9.2	-2.2	11
10/12/87	6.5	14.0	6.8	4.8	11
10/19/87	4.2	6.5	6.8	-0.3	11
10/26/87	4.5	4.2	6.8	-2.6	11
11/02/87	7.2	4.5	4.0	-2.3	11
11/09/87	2.9	7.2	4.0	3.2	11
11/16/87	1.9	2.9	4.0	-1.1	11

Date	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
01/07/85	1	1.01	1	1.01
01/14/85	1	0.75	2	1.76
01/21/85	1	0.31	3	2.07
01/28/85	0	0.31	3	2.39
02/04/85	1	0.97	4	3.36
02/11/85	1	1.00	5	4.36
02/18/85	3	3.00	8	7.36
02/25/85	1	0.44	9	7.80
03/04/85	0	0.67	9	8.48
03/11/85	0	0.78	9	9.26
03/18/85	0	0.40	9	9.66
03/25/85	0	0.40	9	10.06
04/01/85	0	0.40	9	10.46
04/08/85	1	0.40	10	10.86
04/15/85	0	0.40	10	11.26
04/22/85	1	0.06	11	11.32
04/29/85	1	0.40	12	11.72
05/06/85	0	0.06	12	11.79
05/13/85	0	0.06	12	11.85
05/20/85	0	0.06	12	11.91
05/27/85	0	0.06	12	11.98
06/03/85	0	0.06	12	12.04
06/10/85	0	0.06	12	12.10
06/17/85	0	0.06	12	12.17
06/24/85	0	0.06	12	12.23
07/01/85	0	0.06	12	12.30
07/08/85	0	0.06	12	12.36
07/15/85	0	0.06	12	12.42
07/22/85	0	0.06	12	12.49
07/29/85	0	0.06	12	12.55
08/05/85	0	0.06	12	12.61
08/12/85	0	0.06	12	12.68
08/19/85	1	0.06	13	12.74
08/26/85	0	0.06	13	12.80
09/02/85	0	0.06	13	12.87
09/09/85	1	0.06	14	12.93
09/16/85	0	0.07	14	13.00
09/23/85	0	0.06	14	13.06
09/30/85	1	0.39	15	13.46
10/07/85	0	0.40	15	13.86
10/14/85	0	0.36	15	14.22
10/21/85	0	0.37	15	14.59
10/28/85	1	0.36	16	14.95
11/04/85	0	0.36	16	15.32
11/11/85	4	4.27	20	19.59
11/18/85	1	0.36	21	19.95
11/25/85	2	2.63	23	22.58
12/02/85	6	6.00	29	28.58
12/09/85	4	5.78	33	34.36
12/16/85	2	1.12	35	35.48

Date	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
12/23/85	1	0.44	36	35.92
12/30/85	2	0.83	38	36.75
01/06/86	0	0.30	38	37.05
01/13/86	1	0.44	39	37.49
01/20/86	0	0.29	39	37.79
01/27/86	1	0.29	40	38.08
02/03/86	0	0.29	40	38.37
02/10/86	2	2.00	42	40.37
02/17/86	1	1.59	43	41.96
02/24/86	4	6.00	47	47.96
03/03/86	1	0.44	48	48.40
03/10/86	0	0.33	48	48.73
03/17/86	1	0.35	49	49.08
03/24/86	1	0.87	50	49.95
03/31/86	0	0.70	50	50.65
04/07/86	1	0.48	51	51.13
04/14/86	0	0.29	51	51.42
04/21/86	0	0.40	51	51.82
04/28/86	0	0.41	51	52.23
05/05/86	0	0.91	51	53.14
05/12/86	0	0.06	51	53.21
05/19/86	0	0.06	51	53.27
05/26/86	0	0.06	51	53.33
06/02/86	0	0.06	51	53.40
06/09/86	0	0.09	51	53.48
06/16/86	0	0.06	51	53.55
06/23/86	0	0.06	51	53.61
06/30/86	0	0.06	51	53.67
07/07/86	0	0.06	51	53.74
07/14/86	0	0.06	51	53.80
07/21/86	0	0.06	51	53.86
07/28/86	0	0.06	51	53.93
08/04/86	0	0.06	51	53.99
08/11/86	0	0.06	51	54.05
08/18/86	0	0.06	51	54.12
08/25/86	0	0.06	51	54.18
09/01/86	0	0.06	51	54.24
09/08/86	0	0.06	51	54.31
09/15/86	0	0.06	51	54.37
09/22/86	0	0.06	51	54.43
09/29/86	0	0.06	51	54.50
10/06/86	0	0.42	51	54.92
10/13/86	0	0.95	51	55.87
10/20/86	0	0.41	51	56.28
10/27/86	0	0.35	51	56.63
11/03/86	0	0.34	51	56.97
11/10/86	0	0.30	51	57.26
11/17/86	1	2.82	52	60.08
11/24/86	0	1.61	52	61.69
12/01/86	0	0.29	52	61.98

Date	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
12/08/86	1	1.01	53	62.99
12/15/86	0	0.29	53	63.29
12/22/86	0	0.29	53	63.58
12/29/86	0	0.30	53	63.88
01/05/87	0	0.29	53	64.17
01/12/87	1	0.30	54	64.47
01/19/87	0	0.29	54	64.76
01/26/87	1	0.29	55	65.05
02/02/87	0	0.29	55	65.35
02/09/87	0	0.31	55	65.66
02/16/87	0	0.29	55	65.95
02/23/87	0	0.30	55	66.25
03/02/87	1	0.29	56	66.55
03/09/87	0	0.29	56	66.84
03/16/87	0	0.29	56	67.13
03/23/87	1	0.00	57	67.13
03/30/87	0	0.00	57	67.13
04/06/87	1	0.31	58	67.44
04/13/87	1	0.30	59	67.74
04/20/87	1	0.30	60	68.04
04/27/87	1	0.30	61	68.34
05/04/87	0	0.06	61	68.40
05/11/87	0	0.06	61	68.47
05/18/87	0	0.06	61	68.53
05/25/87	0	0.06	61	68.59
06/01/87	0	0.06	61	68.66
06/08/87	0	0.06	61	68.72
06/15/87	0	0.06	61	68.78
06/22/87	0	0.06	61	68.85
06/29/87	0	0.06	61	68.91
07/06/87	0	0.06	61	68.97
07/13/87	0	0.06	61	69.04
07/20/87	0	0.06	61	69.10
07/27/87	0	0.06	61	69.16
08/03/87	0	0.06	61	69.23
08/10/87	0	0.06	61	69.29
08/17/87	0	0.06	61	69.35
08/24/87	1	0.06	62	69.42
08/31/87	0	0.06	62	69.48
09/07/87	0	0.06	62	69.54
09/14/87	0	0.06	62	69.61
09/21/87	0	0.06	62	69.67
09/28/87	0	0.06	62	69.73
10/05/87	0	0.06	62	69.80
10/12/87	0	0.00	62	69.80
10/19/87	0	0.30	62	70.09
10/26/87	0	0.30	62	70.39
11/02/87	0	0.30	62	70.69
11/09/87	0	0.29	62	70.98
11/16/87	0	0.29	62	71.28

APPENDIX B. Sample Model Input Data, Model B.

Date	Temp,7d Average	Temp,n-7d Average	max(t7-t1)	max abs(t7-t1)	max dT,7d	max dT,n-7d 1 week lag
01/07/85	-2.3	-26.3	-0.7	32.0	-8.0	-4.2
01/14/85	-8.8	-2.3	-19.5	19.5	-5.0	-8.0
01/21/85	-9.9	-8.8	-19.7	19.7	-12.4	-5.0
01/28/85	-4.8	-9.9	-7.1	17.9	-4.5	-12.4
02/04/85	-23.4	-4.8	-24.9	24.9	-12.6	-4.5
02/11/85	-23.0	-23.4	-10.1	10.1	-7.4	-12.6
02/18/85	-6.2	-23.0	6.4	25.5	-4.4	-7.4
02/25/85	-3.8	-6.2	-9.8	11.1	-7.0	-4.4
03/04/85	-4.7	-3.8	-10.9	10.9	-10.5	-7.0
03/11/85	-3.1	-4.7	-14.8	15.2	-2.4	-10.5
03/18/85	3.5	-3.1	0.5	9.7	-2.0	-2.4
03/25/85	1.3	3.5	-7.3	7.3	-2.9	-2.0
04/01/85	2.5	1.3	-2.7	9.9	-1.6	-2.9
04/08/85	3.4	2.5	-6.8	6.8	-5.0	-1.6
04/15/85	7.6	3.4	-5.7	10.2	-7.1	-5.0
04/22/85	4.1	7.6	-7.8	7.8	-3.9	-7.1
04/29/85	4.7	4.1	-4.7	7.8	-1.1	-3.9
05/06/85	12.1	4.7	2.2	11.9	-6.2	-1.1
05/13/85	11.2	12.1	-5.2	5.2	-4.4	-6.2
05/20/85	15.5	11.2	-1.0	11.8	-3.4	-4.4
05/27/85	14.4	15.5	-7.8	7.8	-4.9	-3.4
06/03/85	11.1	14.4	-7.5	7.5	-3.7	-4.9
06/10/85	13.4	11.1	-4.3	6.8	-3.2	-3.7
06/17/85	13.7	13.4	-3.6	3.6	-4.5	-3.2
06/24/85	14.3	13.7	-10.1	10.1	-4.9	-4.5
07/01/85	16.7	14.3	-8.3	8.7	-3.7	-4.9
07/08/85	21.6	16.7	-0.3	8.2	-2.2	-3.7
07/15/85	18.8	21.6	-6.4	6.4	-5.0	-2.2
07/22/85	17.9	18.8	-6.4	6.4	-5.4	-5.0
07/29/85	16.7	17.9	-6.4	6.4	-2.9	-5.4
08/05/85	21.0	16.7	-1.3	6.6	-3.8	-2.9
08/12/85	13.0	21.0	-12.9	12.9	-6.9	-3.8
08/19/85	15.2	13.0	-3.6	6.0	-0.7	-6.9
08/26/85	14.5	15.2	-4.3	4.3	-6.1	-0.7
09/02/85	13.2	14.5	-8.4	8.4	-2.5	-6.1
09/09/85	7.7	13.2	-10.7	10.7	-4.8	-2.5
09/16/85	12.2	7.7	-1.0	11.4	-1.8	-4.8
09/23/85	4.7	12.2	-11.1	11.1	-6.7	-1.8
09/30/85	5.6	4.7	-6.6	6.6	-5.5	-6.7
10/07/85	5.6	5.6	-11.7	11.7	-6.1	-5.5
10/14/85	4.1	5.6	-12.4	12.4	-2.8	-6.1
10/21/85	6.4	4.1	-3.8	7.6	-2.9	-2.8

Date	Temp,7d Average	Temp,n-7d Average	max(t7-t1)	max abs(t7-t1)	max dT,7d	max dT,n-7d 1 week lag
10/28/85	3.9	6.4	-8.1	8.1	-3.7	-2.9
11/04/85	0.8	3.9	-6.3	6.3	-1.4	-3.7
11/11/85	-10.3	0.8	-15.9	15.9	-5.2	-1.4
11/18/85	-6.9	-10.3	-11.0	15.5	-10.7	-5.2
11/25/85	-22.4	-6.9	-22.7	22.7	-6.3	-10.7
12/02/85	-21.8	-22.4	-10.0	12.2	-2.7	-6.3
12/09/85	-7.5	-21.8	6.2	19.1	-5.3	-2.7
12/16/85	-8.3	-7.5	-16.6	16.6	-9.4	-5.3
12/23/85	2.8	-8.3	4.0	21.4	-3.8	-9.4
12/30/85	-2.2	2.8	-15.1	15.1	-5.7	-3.8
01/06/86	-7.7	-2.2	-9.0	9.0	-2.9	-5.7
01/13/86	-0.1	-7.7	1.8	13.4	-2.8	-2.9
01/20/86	-3.2	-0.1	-11.4	11.4	-4.8	-2.8
01/27/86	-7.1	-3.2	-11.5	11.5	-3.8	-4.8
02/03/86	-6.7	-7.1	-6.0	9.4	-6.7	-3.8
02/10/86	-11.9	-6.7	-12.4	12.4	-5.6	-6.7
02/17/86	-16.6	-11.9	-16.6	16.6	-8.4	-5.6
02/24/86	-17.7	-16.6	-15.2	25.3	0.0	-8.4
03/03/86	5.0	-17.7	-2.0	30.8	-3.5	0.0
03/10/86	-3.8	5.0	-19.3	19.3	-10.0	-3.5
03/17/86	0.7	-3.8	0.1	10.7	-2.0	-10.0
03/24/86	2.8	0.7	-5.4	6.7	-6.7	-2.0
03/31/86	5.4	2.8	-5.3	10.3	-2.5	-6.7
04/07/86	6.6	5.4	-6.3	6.6	-1.0	-2.5
04/14/86	-1.1	6.6	-18.3	18.3	-9.2	-1.0
04/21/86	7.8	-1.1	-1.5	17.8	-2.7	-9.2
04/28/86	6.1	7.8	-3.8	3.8	-5.0	-2.7
05/05/86	5.4	6.1	-6.5	6.5	-4.9	-5.0
05/12/86	9.2	5.4	-1.2	10.6	-1.5	-4.9
05/19/86	10.2	9.2	-5.3	5.8	-5.6	-1.5
05/26/86	16.1	10.2	-1.9	16.6	-10.4	-5.6
06/02/86	21.8	16.1	-3.4	15.5	-6.1	-10.4
06/09/86	15.9	21.8	-10.7	10.7	-5.2	-6.1
06/16/86	13.9	15.9	-7.6	7.6	-8.0	-5.2
06/23/86	17.4	13.9	-0.6	6.2	-5.1	-8.0
06/30/86	15.5	17.4	-4.1	4.1	-5.2	-5.1
07/07/86	15.8	15.5	-4.0	5.5	-4.0	-5.2
07/14/86	15.9	15.8	-4.5	4.8	-2.6	-4.0
07/21/86	16.4	15.9	-3.6	5.8	-1.4	-2.6
07/28/86	16.0	16.4	-7.2	7.2	-6.6	-1.4
08/04/86	17.5	16.0	-2.1	7.1	-2.8	-6.6
08/11/86	17.5	17.5	-6.3	6.3	-4.0	-2.8

Date	Temp,7d Average	Temp,n-7d Average	max(t7-t1)	max abs(t7-t1)	max dT,7d	max dT,n-7d 1 week lag
08/18/86	17.4	17.5	-3.5	3.5	-3.1	-4.0
08/25/86	14.5	17.4	-7.7	7.7	-6.7	-3.1
09/01/86	17.2	14.5	0.6	4.7	-3.2	-6.7
09/08/86	11.1	17.2	-9.8	9.8	-4.0	-3.2
09/15/86	7.4	11.1	-6.6	6.6	-5.2	-4.0
09/22/86	8.3	7.4	-4.5	4.8	-0.3	-5.2
09/29/86	8.7	8.3	-6.1	6.2	-5.2	-0.3
10/06/86	7.5	8.7	-7.9	8.1	-3.7	-5.2
10/13/86	6.9	7.5	-7.1	7.1	-9.3	-3.7
10/20/86	10.9	6.9	-2.4	9.3	-2.7	-9.3
10/27/86	9.3	10.9	-4.7	4.7	-2.9	-2.7
11/03/86	0.1	9.3	-14.0	14.0	-6.8	-2.9
11/10/86	-7.7	0.1	-25.5	25.5	-10.5	-6.8
11/17/86	-13.6	-7.7	-19.9	19.9	-7.0	-10.5
11/24/86	-10.3	-13.6	-8.2	21.2	-3.2	-7.0
12/01/86	-4.0	-10.3	-7.2	16.4	-6.9	-3.2
12/08/86	-8.2	-4.0	-10.9	10.9	-9.4	-6.9
12/15/86	-3.8	-8.2	-2.9	12.9	-4.4	-9.4
12/22/86	-3.8	-3.8	-3.9	3.9	-3.6	-4.4
12/29/86	-1.9	-3.8	-0.3	4.5	-2.7	-3.6
01/05/87	-3.8	-1.9	-6.1	6.1	-5.3	-2.7
01/12/87	-2.8	-3.8	-3.7	9.6	-3.2	-5.3
01/19/87	-6.1	-2.8	-8.5	8.5	-7.9	-3.2
01/26/87	-5.2	-6.1	-10.5	13.7	-4.0	-7.9
02/02/87	-4.6	-5.2	-2.7	5.4	-3.9	-4.0
02/09/87	-0.1	-4.6	-2.3	10.0	-2.5	-3.9
02/16/87	-3.1	-0.1	-9.9	9.9	-3.1	-2.5
02/23/87	-1.1	-3.1	-6.0	7.2	-3.9	-3.1
03/02/87	-10.3	-1.1	-11.8	11.8	-3.8	-3.9
03/09/87	-7.5	-10.3	-4.3	10.7	-6.2	-3.8
03/16/87	-6.6	-7.5	-11.9	11.9	-1.2	-6.2
03/23/87	-1.6	-6.6	-3.8	10.5	-2.4	-1.2
03/30/87	-2.1	-1.6	-10.9	10.9	-11.4	-2.4
04/06/87	6.7	-2.1	0.7	21.5	-3.7	-11.4
04/13/87	4.7	6.7	-7.2	7.2	-7.1	-3.7
04/20/87	6.8	4.7	-3.0	8.5	-4.0	-7.1
04/27/87	9.4	6.8	-3.5	6.8	-3.5	-4.0
05/04/87	14.2	9.4	-5.5	9.8	-7.5	-3.5
05/11/87	14.4	14.2	-4.1	6.8	-4.2	-7.5
05/18/87	11.3	14.4	-11.2	11.2	-4.5	-4.2
05/25/87	9.2	11.3	-11.7	15.1	-8.1	-4.5
06/01/87	14.7	9.2	-3.6	12.7	-3.8	-8.1

Date	max abs (dT)	max abs dT,n-7d 1 week lag	Wat temp,7d Average	T7(air-wat) Average	T7(a-w),n-7d Average	150,n-1wk 1 week lag
01/07/85	11.8	6.7	9.7	-12.0	-35.8	3
01/14/85	13.1	11.8	9.7	-18.5	-12.0	1
01/21/85	12.4	13.1	9.7	-19.6	-18.5	1
01/28/85	5.2	12.4	9.7	-14.5	-19.6	1
02/04/85	12.6	5.2	9.8	-33.2	-14.5	0
02/11/85	7.4	12.6	9.9	-32.9	-33.2	1
02/18/85	7.0	7.4	9.9	-16.1	-32.9	1
02/25/85	7.0	7.0	9.9	-13.7	-16.1	3
03/04/85	10.5	7.0	10.0	-14.7	-13.7	1
03/11/85	6.4	10.5	10.1	-13.2	-14.7	0
03/18/85	2.3	6.4	10.1	-6.6	-13.2	0
03/25/85	3.3	2.3	10.1	-8.8	-6.6	0
04/01/85	3.2	3.3	10.1	-7.6	-8.8	0
04/08/85	7.3	3.2	10.1	-6.7	-7.6	0
04/15/85	7.1	7.3	10.1	-2.5	-6.7	1
04/22/85	4.9	7.1	10.1	-6.0	-2.5	0
04/29/85	2.9	4.9	10.1	-5.4	-6.0	1
05/06/85	6.2	2.9	15.1	-3.0	-5.4	1
05/13/85	4.4	6.2	15.1	-3.9	-3.0	0
05/20/85	3.4	4.4	15.1	0.4	-3.9	0
05/27/85	4.9	3.4	15.1	-0.7	0.4	0
06/03/85	3.7	4.9	15.8	-4.7	-0.7	0
06/10/85	6.0	3.7	16.7	-3.3	-4.7	0
06/17/85	4.5	6.0	16.7	-3.0	-3.3	0
06/24/85	4.9	4.5	16.7	-2.4	-3.0	0
07/01/85	5.6	4.9	16.7	0.0	-2.4	0
07/08/85	2.2	5.6	21.9	-0.3	0.0	0
07/15/85	5.0	2.2	21.9	-3.1	-0.3	0
07/22/85	5.4	5.0	21.9	-4.0	-3.1	0
07/29/85	2.9	5.4	21.9	-5.2	-4.0	0
08/05/85	3.8	2.9	19.3	1.7	-5.2	0
08/12/85	6.9	3.8	18.3	-5.3	1.7	0
08/19/85	3.2	6.9	18.3	-3.1	-5.3	0
08/26/85	6.1	3.2	18.3	-3.8	-3.1	1
09/02/85	2.5	6.1	16.3	-3.1	-3.8	0
09/09/85	4.8	2.5	11.2	-3.5	-3.1	0
09/16/85	2.0	4.8	11.2	1.0	-3.5	1
09/23/85	6.7	2.0	11.2	-6.5	1.0	0
09/30/85	5.7	6.7	11.2	-5.6	-6.5	0
10/07/85	6.1	5.7	6.0	-0.4	-5.6	1
10/14/85	8.7	6.1	6.0	-1.9	-0.4	0
10/21/85	4.8	8.7	6.0	0.4	-1.9	0

Date	max abs (dT)	max abs dT,n-7d 1 week lag	Wat temp,7d Average	T7(air-wat) Average	T7(a-w),n-7d Average	150,n-1wk 1 week lag
10/28/85	3.7	4.8	6.0	-2.1	0.4	0
11/04/85	1.7	3.7	5.9	-5.2	-2.1	1
11/11/85	5.2	1.7	5.9	-16.2	-5.2	0
11/18/85	10.7	5.2	5.9	-12.8	-16.2	4
11/25/85	7.0	10.7	5.9	-28.3	-12.8	1
12/02/85	7.8	7.0	5.8	-27.6	-28.3	2
12/09/85	10.0	7.8	5.6	-13.1	-27.6	6
12/16/85	11.7	10.0	5.6	-13.9	-13.1	4
12/23/85	4.1	11.7	5.6	-2.8	-13.9	2
12/30/85	6.0	4.1	5.6	-7.8	-2.8	1
01/06/86	2.9	6.0	5.5	-13.2	-7.8	2
01/13/86	6.8	2.9	5.5	-5.6	-13.2	0
01/20/86	4.8	6.8	5.5	-8.7	-5.6	1
01/27/86	6.1	4.8	5.5	-12.6	-8.7	0
02/03/86	6.7	6.1	5.5	-12.2	-12.6	1
02/10/86	5.6	6.7	5.4	-17.3	-12.2	0
02/17/86	8.4	5.6	5.4	-22.0	-17.3	2
02/24/86	11.9	8.4	5.4	-23.1	-22.0	1
03/03/86	6.6	11.9	5.1	-0.1	-23.1	4
03/10/86	10.0	6.6	4.8	-8.6	-0.1	1
03/17/86	2.0	10.0	4.8	-4.1	-8.6	0
03/24/86	6.7	2.0	4.8	-2.0	-4.1	1
03/31/86	8.6	6.7	4.8	0.6	-2.0	1
04/07/86	4.5	8.6	5.1	1.5	0.6	0
04/14/86	9.2	4.5	5.1	-6.2	1.5	1
04/21/86	6.4	9.2	5.1	2.7	-6.2	0
04/28/86	5.0	6.4	5.1	1.0	2.7	0
05/05/86	8.1	5.0	10.4	-5.0	1.0	0
05/12/86	3.6	8.1	12.5	-3.3	-5.0	0
05/19/86	5.6	3.6	12.5	-2.3	-3.3	0
05/26/86	10.4	5.6	12.5	3.6	-2.3	0
06/02/86	6.1	10.4	14.0	7.7	3.6	0
06/09/86	5.2	6.1	17.9	-2.0	7.7	0
06/16/86	8.0	5.2	17.9	-4.0	-2.0	0
06/23/86	5.8	8.0	17.9	-0.5	-4.0	0
06/30/86	5.2	5.8	17.9	-2.4	-0.5	0
07/07/86	4.0	5.2	16.9	-1.1	-2.4	0
07/14/86	3.4	4.0	16.9	-1.0	-1.1	0
07/21/86	2.4	3.4	16.9	-0.5	-1.0	0
07/28/86	6.6	2.4	16.9	-0.9	-0.5	0
08/04/86	2.8	6.6	18.3	-0.7	-0.9	0
08/11/86	4.0	2.8	19.3	-1.8	-0.7	0

Date	max abs (dT)	max abs dT,n-7d 1 week lag	Wat temp,7d Average	T7(air-wat) Average	T7(a-w),n-7d Average	150,n-1wk 1 week lag
08/18/86	4.0	4.0	19.3	-1.9	-1.8	0
08/25/86	6.7	4.0	19.3	-4.8	-1.9	0
09/01/86	3.4	6.7	19.3	-2.1	-4.8	0
09/08/86	4.0	3.4	12.2	-1.1	-2.1	0
09/15/86	5.2	4.0	12.2	-4.8	-1.1	0
09/22/86	3.8	5.2	12.2	-3.9	-4.8	0
09/29/86	5.2	3.8	12.2	-3.5	-3.9	0
10/06/86	9.0	5.2	7.8	-0.3	-3.5	0
10/13/86	9.3	9.0	7.8	-0.9	-0.3	0
10/20/86	3.9	9.3	7.8	3.1	-0.9	0
10/27/86	2.9	3.9	7.8	1.5	3.1	0
11/03/86	10.4	2.9	6.5	-6.4	1.5	0
11/10/86	10.5	10.4	4.8	-12.5	-6.4	0
11/17/86	9.6	10.5	4.8	-18.4	-12.5	0
11/24/86	10.2	9.6	4.8	-15.1	-18.4	1
12/01/86	6.9	10.2	4.8	-8.8	-15.1	0
12/08/86	9.4	6.9	5.1	-13.3	-8.8	0
12/15/86	9.6	9.4	5.1	-8.9	-13.3	1
12/22/86	4.3	9.6	5.1	-8.9	-8.9	0
12/29/86	3.6	4.3	5.1	-7.0	-8.9	0
01/05/87	5.3	3.6	5.2	-9.1	-7.0	0
01/12/87	6.2	5.3	5.3	-8.1	-9.1	0
01/19/87	7.9	6.2	5.3	-11.4	-8.1	1
01/26/87	4.4	7.9	5.3	-10.5	-11.4	0
02/02/87	3.9	4.4	5.3	-10.0	-10.5	1
02/09/87	4.5	3.9	5.4	-5.5	-10.0	0
02/16/87	3.1	4.5	5.4	-8.5	-5.5	0
02/23/87	3.9	3.1	5.4	-6.5	-8.5	0
03/02/87	4.0	3.9	5.4	-15.7	-6.5	0
03/09/87	9.9	4.0	5.3	-12.8	-15.7	1
03/16/87	3.5	9.9	5.3	-11.9	-12.8	0
03/23/87	6.8	3.5	5.3	-6.9	-11.9	0
03/30/87	11.4	6.8	5.3	-7.4	-6.9	1
04/06/87	5.5	11.4	3.9	2.8	-7.4	0
04/13/87	7.1	5.5	3.9	0.8	2.8	1
04/20/87	4.0	7.1	3.9	2.9	0.8	1
04/27/87	7.9	4.0	3.9	5.5	2.9	1
05/04/87	7.5	7.9	9.8	4.3	5.5	1
05/11/87	6.0	7.5	14.3	0.1	4.3	0
05/18/87	6.7	6.0	14.3	-3.0	0.1	0
05/25/87	8.1	6.7	14.3	-5.1	-3.0	0
06/01/87	3.8	8.1	14.3	0.4	-5.1	0

Date	150,n-1yr 1 yr prev.	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
01/07/85	15	1	5.57	1	5.57
01/14/85	15	1	1.02	2	6.59
01/21/85	16	1	1.07	3	7.66
01/28/85	16	0	0.00	3	7.66
02/04/85	17	1	0.99	4	8.65
02/11/85	18	1	0.98	5	9.62
02/18/85	21	3	3.18	8	12.80
02/25/85	22	1	1.90	9	14.71
03/04/85	22	0	0.41	9	15.12
03/11/85	21	0	0.88	9	16.00
03/18/85	21	0	0.56	9	16.56
03/25/85	21	0	0.33	9	16.89
04/01/85	21	0	0.71	9	17.60
04/08/85	21	1	0.27	10	17.87
04/15/85	21	0	0.24	10	18.11
04/22/85	22	1	0.76	11	18.87
04/29/85	23	1	1.07	12	19.94
05/06/85	23	0	0.00	12	19.94
05/13/85	23	0	0.02	12	19.95
05/20/85	23	0	0.07	12	20.03
05/27/85	23	0	0.00	12	20.03
06/03/85	23	0	0.11	12	20.14
06/10/85	23	0	0.03	12	20.17
06/17/85	22	0	0.00	12	20.17
06/24/85	22	0	0.00	12	20.17
07/01/85	22	0	0.00	12	20.17
07/08/85	22	0	0.00	12	20.17
07/15/85	22	0	0.00	12	20.17
07/22/85	22	0	0.00	12	20.17
07/29/85	22	0	0.00	12	20.17
08/05/85	22	0	0.00	12	20.17
08/12/85	21	0	0.00	12	20.17
08/19/85	22	1	0.08	13	20.25
08/26/85	22	0	0.00	13	20.25
09/02/85	22	0	0.06	13	20.31
09/09/85	23	1	0.92	14	21.22
09/16/85	23	0	0.96	14	22.18
09/23/85	23	0	0.16	14	22.34
09/30/85	24	1	0.91	15	23.25
10/07/85	24	0	1.07	15	24.32
10/14/85	24	0	1.31	15	25.63
10/21/85	23	0	0.56	15	26.19

Date	150,n-1yr 1 yr prev.	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
10/28/85	24	1	0.72	16	26.91
11/04/85	24	0	1.13	16	28.04
11/11/85	27	4	2.09	20	30.12
11/18/85	27	1	0.00	21	30.12
11/25/85	29	2	1.50	23	31.62
12/02/85	34	6	3.18	29	34.81
12/09/85	38	4	0.00	33	34.81
12/16/85	40	2	0.00	35	34.81
12/23/85	39	1	1.05	36	35.86
12/30/85	40	2	2.04	38	37.90
01/06/86	37	0	0.08	38	37.98
01/13/86	38	1	0.86	39	38.84
01/20/86	36	0	0.00	39	38.84
01/27/86	37	1	0.85	40	39.69
02/03/86	36	0	0.00	40	39.69
02/10/86	37	2	1.99	42	41.67
02/17/86	35	1	1.01	43	42.68
02/24/86	38	4	4.08	47	46.76
03/03/86	39	1	1.00	48	47.76
03/10/86	39	0	0.05	48	47.81
03/17/86	40	1	0.96	49	48.77
03/24/86	41	1	0.00	50	48.77
03/31/86	41	0	0.70	50	49.47
04/07/86	41	1	0.94	51	50.41
04/14/86	41	0	0.00	51	50.41
04/21/86	40	0	0.50	51	50.91
04/28/86	40	0	0.18	51	51.09
05/05/86	39	0	0.05	51	51.13
05/12/86	39	0	0.00	51	51.13
05/19/86	39	0	0.03	51	51.16
05/26/86	39	0	0.00	51	51.16
06/02/86	39	0	0.69	51	51.84
06/09/86	39	0	0.25	51	52.10
06/16/86	39	0	0.40	51	52.50
06/23/86	39	0	0.59	51	53.09
06/30/86	39	0	0.02	51	53.11
07/07/86	39	0	0.01	51	53.13
07/14/86	39	0	0.00	51	53.13
07/21/86	39	0	0.00	51	53.13
07/28/86	39	0	0.00	51	53.13
08/04/86	39	0	0.00	51	53.13
08/11/86	39	0	0.00	51	53.13

Date	150,n-1yr 1 yr prev.	Actual Breaks	Model Output	Actual Cumulative	Model Cumulative
08/18/86	38	0	0.00	51	53.13
08/25/86	38	0	0.00	51	53.13
09/01/86	38	0	0.00	51	53.13
09/08/86	38	0	0.04	51	53.16
09/15/86	37	0	0.04	51	53.20
09/22/86	37	0	0.05	51	53.25
09/29/86	37	0	0.04	51	53.29
10/06/86	36	0	0.10	51	53.39
10/13/86	36	0	0.00	51	53.39
10/20/86	36	0	0.00	51	53.39
10/27/86	35	0	0.10	51	53.50
11/03/86	35	0	0.00	51	53.50
11/10/86	34	0	0.00	51	53.50
11/17/86	32	1	0.95	52	54.45
11/24/86	31	0	0.00	52	54.45
12/01/86	26	0	0.09	52	54.54
12/08/86	22	1	0.96	53	55.50
12/15/86	19	0	0.45	53	55.95
12/22/86	17	0	0.40	53	56.35
12/29/86	16	0	0.04	53	56.39
01/05/87	15	0	0.00	53	56.39
01/12/87	15	1	0.27	54	56.66
01/19/87	15	0	0.26	54	56.93
01/26/87	15	1	0.73	55	57.66
02/02/87	15	0	0.11	55	57.77
02/09/87	13	0	0.19	55	57.96
02/16/87	12	0	0.05	55	58.01
02/23/87	9	0	0.22	55	58.22
03/02/87	8	1	1.10	56	59.32
03/09/87	8	0	0.46	56	59.78
03/16/87	7	0	0.03	56	59.82
03/23/87	8	1	0.98	57	60.80
03/30/87	7	0	0.28	57	61.08
04/06/87	7	1	0.00	58	61.08
04/13/87	8	1	0.84	59	61.92
04/20/87	9	1	0.75	60	62.67
04/27/87	10	1	1.00	61	63.67
05/04/87	10	0	0.00	61	63.67
05/11/87	10	0	0.01	61	63.69
05/18/87	10	0	0.03	61	63.71
05/25/87	10	0	0.10	61	63.82
06/01/87	10	0	0.02	61	63.84